



Kofinanziert von der
Europäischen Union

 **Ernst-Abbe-Hochschule Jena**
University of Applied Sciences

Manja Ellrich, Dorothea Gläßer, Christopher Holl, Annalena
Schröder und Dr. Sebastian Schulz

ABSCHLUSSBERICHT

Evaluation des Sozialpädagogischen Teamteachings an
Thüringer Schulen:

Eine Studie zur Erforschung der Wirk- und Gelingensfaktoren
2023 – 2025

Projektleitung:

Prof. Dr. Nicole S. Harth & Prof. Dr. Kristin Mitte



Impressum

Projektleitung | Prof. Dr. Nicole S. Harth & Prof. Dr. Kristin Mitte

Fachbereich Sozialwesen

Ernst-Abbe-Hochschule Jena

Carl-Zeiss-Promenade 2

07745 Jena

Wissenschaftliche Mitarbeiter:innen | Manja Ellrich, Dorothea Gläßer, Christopher Holl, Annalena Schröder, Dr. Sebastian Schulz

Kontakt | nicole.harth@eah-jena.de

Zitiervorschlag | Harth, N. S., Mitte, K., Ellrich, M., Gläßer, D., Holl, C., Schröder, A., Schulz, S. (2025). *Evaluation des Sozialpädagogischen Teamteachings an Thüringer Schulen: Eine Studie zur Erforschung der Wirk- und Gelingens-Faktoren*. Abschlussbericht der Projektphase 2023-2025. Ernst-Abbe-Hochschule Jena.

Danksagung

Wir bedanken uns bei allen Schüler:innen, die uns ihr Vertrauen geschenkt haben, ihre Alltagserfahrungen mit uns geteilt haben und somit diese Evaluation ermöglicht haben sowie deren Eltern für die Unterstützung unserer Forschungsarbeit.

Wir danken allen beteiligten Projektpartner:innen für die wertvolle Zusammenarbeit, insbesondere den Teamteachenden, Lehrkräften und Regionalteamleitungen, die uns (teilweise über die eigene Teilnahme an der Evaluation hinaus) bei den Erhebungen unterstützt haben. Außerdem danken wir unseren studentischen Assistentinnen und (ehemaligen) Kolleg:innen, die uns in den unterschiedlichen Phasen des Projektes unterstützt haben: Inya Bergmann, Marco Böhm, Sarah Ehlers, Laura Golle, Josefine Hapke, Theodora Laoumtzis, Juliane Salzer, Lucy Schmidt, Cora Schüll, Karina Völk, Jonas Wiemann sowie Sarah Boelter, Petra Gordalla, Hannah Ingwersen, Dr. Maya Machunsky, und Laura McCullagh.

Inhaltsverzeichnis

Impressum	2
Abkürzungsverzeichnis.....	6
Abbildungsverzeichnis	7
Tabellenverzeichnis.....	8
TEIL I – EINLEITUNG.....	9
1.1 Das Projekt „Evaluierung des Sozialpädagogischen Teamteachings“	9
1.2 Theoretischer Hintergrund	10
1.2.1 Internationaler Forschungsstand zur Wirksamkeit des Teamteachings	11
1.2.2 Gelingensbedingungen.....	12
1.3 Methodisches Vorgehen	16
1.3.1 Auswahl der Interventions- und Kontrollgruppen	16
1.3.2 Aufteilung in <i>intention to treat</i> und <i>per protocol</i>	17
1.4 Genehmigung der Erhebungen	18
1.5 Forschungsdatenmanagement	19
1.6 Inklusion, Diversität und Nachhaltigkeit	19
TEIL II – SCHÜLER:INNEN-BEFRAGUNG	22
2.1 Durchführung und Ablauf der Erhebung.....	22
2.1.1 Durchführung der Erhebung.....	22
2.1.2 Stichprobenbeschreibung	23
2.1.3 Rücklauf- und Teilnahmequoten.....	25
2.2 Entwicklung der Fragebögen	26
2.2.1 Theoretischer Hintergrund der Schüler:innen Fragebögen	26
2.2.2 Verwendete Skalen	27
2.3 Vergleichbarkeit zwischen Kontroll- und STT-Gruppe hinsichtlich zentraler soziodemographischer Variablen	36
2.4 Methodik der Datenanalyse	41
2.5 Ergebnisse der Schüler:innenfragebögen.....	43
2.5.1 Skalen zum STT und zu Lehrpersonen	43
2.5.2 Bedürfniserfüllung, Wohlbefinden und Emotionsregulation	45

2.5.3	Schulerfolg	48
2.5.4	Soziales Klassengefüge	54
2.6	Zusammenfassung: Ergebnisse der Schüler:innenbefragung.....	56
TEIL III – ÜBERSICHTEN DER NOTEN UND FEHLTAGE		58
3.1	Durchführung der Erhebung	58
3.2	Methodik der Datenanalyse	58
3.3	Ergebnisse	58
3.3.1	Entschuldigte und unentschuldigte Fehltage.....	58
3.3.1	Klassenlistendaten: Schulnoten	60
3.4	Zusammenfassung: Ergebnisse der ÜNuF-Erhebung.....	61
TEIL IV –BEFRAGUNG DER TEAMTEACHENDEN UND KLASSENLEHRKRÄFTE.....		62
4.1	Durchführung der Erhebung	62
4.2	Stichprobenbeschreibung.....	62
4.3	Verwendete Skalen	63
4.3.1	Stressbelastung	64
4.3.2	Arbeitszufriedenheit	64
4.3.3	Identifikation mit der Schule	64
4.3.4	Commitment gegenüber dem STT	65
4.3.5	Kooperation im Tandem	65
4.4	Methodik der Datenanalyse	66
4.5	Ergebnisse	66
4.5.1	Stressbelastung	67
4.5.2	Arbeitszufriedenheit	67
4.5.3	Identifikation mit der Schule	67
4.5.4	Commitment gegenüber dem STT	68
4.5.5	Kooperation im Tandem	68
4.6	Zusammenfassung: Ergebnisse der LK- und TT-Fragebögen	68
TEIL V – WEITERFÜHRENDE ANALYSEN.....		69
5.1	Einleitung	69

5.2	Analyse der T3-Werte in Abhängigkeit der Dauer an STT.....	69
5.3	Erklärungsansätze	71
5.3.1	Statistische Artefakte & Stichprobencharakteristika	71
5.3.2	Selektionshypothese.....	72
5.3.3	Stigmatisierungshypothese	73
5.3.4	Autonomiehypothese	76
5.4	Zusammenfassung: Weiterführende Analysen	78
TEIL VI – SCHLUSS.....		80
6.1	Zusammenfassung.....	80
6.2	Limitationen und offene Fragen.....	81
6.3	Empfehlungen	83
6.3.1	Empfehlung 1: Stärkerer Fokus auf Autonomie	83
6.3.2	Empfehlung 2: Intensität und Dauer des STT zielgruppengerechter gestalten 84	
6.3.3	Empfehlung 3: Rollenklarheit fördern	84
6.3.4	Empfehlung 4: Reflexion der Wirkebene	85
6.3.5	Empfehlung 5: Vulnerabilität der Schüler:innen im Fokus	85
6.4	Abschlussbewertung	86
6.5	Das Wichtigste auf einen Blick	88
6.6	Empfehlungen auf einen Blick	89
Literaturverzeichnis.....		90
ANHANG.....		101
Anhang 1: Publikationsverzeichnis		101
Vorträge.....		101
Publikationen.....		102
Anhang 2: Deskriptive Werte relevanter Variablen in der Schüler:innen-Stichprobe ...		103

Abkürzungsverzeichnis

ANOVA	Varianzanalyse
EAH	Ernst-Abbe-Hochschule Jena
ESF	Europäischer Sozialfonds
FB	Fragebogen
ITT	intention to treat
KSB	Kindersprachbrücke Jena e.V.
LK	Lehrkraft/Lehrkräfte
MZP	Messzeitpunkt
NEPS	Nationales Bildungspanel aus der Bildungsberichterstattung
PP	per protocol
SDT	Self-Determination Theory nach Ryan und Deci (2000)
SES	Sozioökonomischer Status
SRQ-A[G]	German Academic Self-Regulation Questionnaire
STP	Student Trust in Principal Skala
STT	Sozialpädagogisches Teamteaching
SuS	Schüler:innen
ThürAbmildSchulVO	Thüringer Abmilderungsverordnung
ThürSchulG	Thüringer Schulgesetz
TMBJS	Thüringer Ministerium für Bildung, Jugend und Sport, ab Dezember 2024 Thüringer Ministerium für Bildung, Wissenschaft und Kultur
TMBWK	Thüringer Ministerium für Bildung, Wissenschaft und Kultur
TT	Teamteachende: Sozialarbeiter:innen oder Sozialpädagog:innen, die gemeinsam mit einer Lehrkraft im STT-Tandem arbeiten
T1, T2, T3	Messzeitpunkte 1, 2 und 3
ÜNuF	Übersicht(en) der Noten und Fehltage

Abbildungsverzeichnis

Abbildung 1 Verteilung der Geschlechterangaben nach ITT.	37
Abbildung 2 Verteilung der Altersangaben zum ersten Erhebungszeitpunkt nach ITT.	37
Abbildung 3 Verteilung des höchsten Bildungsabschlusses der Eltern nach ITT.	38
Abbildung 4 Verteilung des höchsten Bildungsabschlusses der Eltern nach PP.....	39
Abbildung 5 Verteilung der Angaben bzgl. der zu Hause zur Verfügung stehenden Lernressourcen nach ITT.	39
Abbildung 6 Verteilung der Bücheranzahl zu Hause nach ITT.	39
Abbildung 7 Anteil SuS mit Migrationshintergrund insgesamt (grau), in Kontrollgruppe (blau) und STT-Gruppe (orange).	40
Abbildung 8 Anteil SuS mit nicht-deutscher Erstsprache in Kontrollgruppe (blau) und STT-Gruppe (orange).	40
Abbildung 9 Zeitlicher Verlauf von Vertrauen der SuS in Teamteachende.	43
Abbildung 10 Zeitlicher Verlauf von Fairness der Teamteachenden.	44
Abbildung 11 Zeitlicher Verlauf der wahrgenommenen Fairness der Klassenlehrkraft.	44
Abbildung 12 Zeitliche Entwicklung des Wohlbefindens im Schulalltag, 1 = negatives und 10 = positives Befinden.....	47
Abbildung 13 Anteil SuS, die angegeben haben, bereits schon einmal geschwänzt zu haben...49	
Abbildung 14 Selbstangabe Absentismusverhalten nach ITT.	50
Abbildung 15 Zeitliche Entwicklung des kognitiven Engagements.	51
Abbildung 16 Zeitliche Entwicklung des verhaltensbezogenen Disengagements.	51
Abbildung 17 Zeitliche Entwicklung des agentischen Engagements.	52
Abbildung 18 Zeitliche Entwicklung des Klassenklimas.	55
Abbildung 19 Zeitliche Entwicklung des negativen Verhaltens innerhalb der Klasse.	56
Abbildung 20 Durchschnittliche Anzahl unentschuldigter Fehltage pro Schuljahr pro Schüler:in gruppiert nach ITT.....	59
Abbildung 21 Durchschnittliche Anzahl entschuldigter Fehltage pro Schuljahr pro Schüler:in gruppiert nach ITT.....	60
Abbildung 22 Zeitliche Entwicklung des Gesamtnotenschnitts (Deutsch, Mathe und Englisch), Unterteilung nach ITT.....	61
Abbildung 23 Anteil SuS, die nach Auffassung der LK aus herausfordernden familiären Verhältnissen kommt, Angaben in Prozent, Differenzierung nach ITT.	74
Abbildung 24 Zustimmung der LK zur Aussage, dass die SuS eine gute Perspektive haben, Differenzierung nach ITT.	74

Tabellenverzeichnis

Tabelle 1 Klassen mit und ohne STT im Laufe der Evaluation.....	17
Tabelle 2 Aufteilung der Klassen in STT- und Kontrollgruppe	18
Tabelle 3 Beschreibung der Schüler:innen-Stichprobe	24
Tabelle 4 Durchschnittliche Klassengrößen, Rücklauf- und Anwesenheitsquoten	25
Tabelle 5 Mittelwerte und Standardabweichungen der wahrgenommenen Autonomieunterstützung durch die Lehrkraft	45
Tabelle 6 Mittelwerte und Standardabweichungen zu den psychologischen Grundbedürfnissen	46
Tabelle 7 Mittelwerte und Standardabweichung für Emotionsregulation	48
Tabelle 8 Mittelwerte und Standardabweichungen zu den drei Facetten von Motivation	53
Tabelle 9 Anteil Schüler:innen mit realistischer Bildungsaspiration "Kein Schulabschluss"	54
Tabelle 10 Unentschuldigte Fehltage.....	59
Tabelle 11 Notenschnitt.....	60
Tabelle 12 Teilnehmende an der Befragung der Tandems und Klassenlehrkräfte	62
Tabelle 13 Demographischer Hintergrund der teilnehmenden Lehrkräfte und Teamteachenden	63
Tabelle 14 Mittelwerte und Standardabweichungen der TT/LK-Erhebung für T1 und T2	66
Tabelle 15 Korrelations- und multiple lineare Regressionsanalysen zwischen der Dauer des STT und abhängigen Variablen, kontrolliert auf den ersten Messzeitpunkt	69
Tabelle 16 Ergebnisse der Mediationsanalysen hinsichtlich des Notenschnittes zu T3	75
Tabelle 17 Ergebnisse der Mediationsanalysen hinsichtlich der intrinsischen Motivation zu T3	75
Tabelle 18 Ergebnisse der Mediationsanalysen hinsichtlich des agentischen Engagements zu T3	75
Tabelle 19 Mediationsanalysen zur Autonomiebedürfniserfüllung	78
Tabelle 20 Deskriptive Werte relevanter Variablen in der Schüler:innen Stichprobe über die Zeit	103

TEIL I – EINLEITUNG

1.1 Das Projekt „Evaluierung des Sozialpädagogischen Teamteachings“

Das Sozialpädagogische Teamteaching (STT) meint die zeitgleiche Zusammenarbeit von Lehrer:in und Sozialpädagog:in (auch Tandem genannt) im Schulalltag einer Klasse und somit die Vernetzung von Jugendhilfe und Schule in der Praxis. Entwickelt wurde das Konzept durch die **Kindersprachbrücke Jena e.V.** (KSB), die Projektgrundlagen sind im Manual „Handbuch Sozialpädagogisches Teamteaching“ (Kindersprachbrücke Jena e. V., 2019) und der aktualisierten Version von 2024 (Kindersprachbrücke Jena e. V., 2024) dargestellt.¹ Aktuell wird das STT an 18 Thüringer Regel- und Gemeinschaftsschulen durchgeführt, wobei jeweils 1-3 Klassen (Klassenstufen 5-9) durch Teamteachende (TT) begleitet werden.

Das Interventionskonzept ist stark praxisorientiert und leitet sich vorrangig aus Erfahrungen und Projekten der KSB ab, nur vereinzelt werden Theoriekonzepte angeschnitten, bspw. die *Selbstbestimmungstheorie* nach Ryan und Deci (2000). Die KSB übernimmt zudem die fachliche Prozessbegleitung, die Projektschulen sind untereinander in fünf Regionalteams vernetzt. Einige Arbeitsgrundlagen werden gemeinsam mit den Tandems verschriftlicht (bspw. Tätigkeitsprofile der TT, Handlungszieltabellen, Handbuch zum STT), die Prozessbegleitung durch die KSB ist eher durch Gesprächsformate geprägt, bspw. Onboarding-Gespräche, Coachings mit Rollenklärungen oder Hospitationen. Einmal im Jahr findet ein „Schulleitungsdinner“ mit allen beteiligten Schulleitungen zum fachlichen Austausch statt. Zwei Mal jährlich treffen sich alle Teamteachenden, die Regionalteamleitungen und das STT-Leitungsteam der KSB in einem zweitägigen Camp zur inhaltlichen Arbeit. Zum Weiterbildungskonzept im STT gehören neben Basisfortbildungen auch Fortbildungen, die sich an den aktuellen Bedarfen der Tandems und Projektschulen orientieren und freiwillig in Anspruch genommen werden können.

Gefördert wird die Implementation des STT an Thüringer Schulen in herausfordernden Lagen aus Mitteln des Europäischen Sozialfonds (ESF), seit 2022 festgehalten im Rahmen der ESF+ Schulförderrichtlinie (Thüringer Ministerium für Bildung, Jugend und Sport, 2022). Ziel des STT ist, soziale Benachteiligungen auszugleichen und individuelle Beeinträchtigungen bei Schüler:innen (SuS) zu überwinden². Durch das Berücksichtigen individueller Bedarfe und Beeinträchtigungen soll das STT dazu beitragen, dass Schulabbruchquoten gesenkt werden. Im

¹ Diese aktualisierte Version konnte jedoch nicht mehr bei der Konzeption der hier berichteten Evaluation berücksichtigt werden, weswegen sich im Folgenden auf das frühere Manual von 2019 bezogen wird.

² Wie in §13 SGB VIII festgehalten.

Manual sind die verschiedenen Interventionsaspekte überblicksartig dargestellt, wobei betont wird, dass das Konzept ständig weiterentwickelt wird (Kindersprachbrücke Jena e. V., 2019, S. 21). Diese Weiterentwicklung geschieht vor allem im Rahmen des regelmäßigen kollegialen Austauschs in den Regionalteams, aber auch Thüringen-übergreifend im Rahmen von Fachtagen, an denen auch Schulleitungen und Lehrkräfte teilnehmen können.

Die Alltagsbegleitung durch Sozialarbeiter:innen soll sich laut Manual nicht nur auf Ebene der Schüler:innen, sondern auch auf Klassen- und Schulebene auswirken. Ausgangspunkt der Zusammenarbeit mit den Lehrkräften bildet das sog. „Kompetenzmodell“, wobei Teamteachende v.a. auf Sozial- und Selbstkompetenzen der Schüler:innen Einfluss nehmen, während Sach- und Methodenkompetenz eher durch die (Tandem-)Lehrkräfte gefördert werden sollen (Kindersprachbrücke Jena e. V., 2019, S. 8). Die Multi-Professionalität, Vernetzung zu außerschulischen Angeboten und Kooperationspartner:innen sowie eine interkulturelle Sensibilisierung sollen sich auch über die Klasse hinaus positiv auf die Schulentwicklung auswirken. Als mögliche Interventionen auf Klassenebene nennt das Manual bspw. Entspannungs- und Konzentrationsübungen, Formen des sozialen Lernens sowie eine verstärkte Vernetzung mit den Erziehungsberechtigten der Schüler:innen (bspw. durch die Mitgestaltung von Elterngesprächen). Welche Übungen und Interventionen konkret angewendet werden, liegt jedoch im Ermessensspielraum der jeweiligen Tandems. Ziel ist dabei, den Unterricht in enger Absprache zwischen Lehrkraft und Teamteacher:in bedarfsorientiert, methodisch abwechslungsreicher und interaktiver zu gestalten.

Im Manual sind zudem Maßnahmen zur Qualitätssicherung festgehalten, bspw. „regelmäßiger Austausch zwischen der Kindersprachbrücke Jena e.V. und den jeweiligen Trägern der Maßnahme“ (Kindersprachbrücke Jena e. V., 2019, S. 19). Über den Rahmen der internen Prozessbegleitung hinaus begleitete das Forschungsprojekt „Evaluierung des Sozialpädagogischen Teamteachings“ an der Ernst-Abbe-Hochschule (EAH) Jena zwischen 2019 und 2022 die Implementation des STT durch eine wissenschaftliche Evaluation (Mitte et al., 2023). **Aufbauend auf die Erkenntnisse der ersten Evaluationsphase wurde zwischen 2023 und 2025 ein weiteres Forschungsprojekt durchgeführt, „Evaluation des Sozialpädagogischen Teamteachings an Thüringer Schulen: eine Studie zur Erforschung der Wirk- und Gelingens-Faktoren (EvaTeamTeach)“, dessen Ergebnis der vorliegende Projektbericht ist.** Im Rahmen des Forschungsprojekts wurden wissenschaftliche Vorträge und Publikationen erstellt, das komplette Publikationsverzeichnis ist dem **ANHANG** zu entnehmen.

1.2 Theoretischer Hintergrund

Das STT baut auf pädagogische Konzepte zum kollaborativen Unterrichten zwischen zwei oder mehr Lehrpersonen auf, die denselben, aber auch verschiedene fachliche Hintergründe

haben können. In der Fachliteratur finden sich v.a. Begriffe wie *collaborative teaching* (bspw. Akiba et al., 2019; Anwar et al., 2021), *co-teaching* (bspw. Härkki et al., 2021) oder *team-teaching* (bspw. Tsybulsky & Muchnik-Rozanov, 2019), die jedoch nicht trennscharf verwendet werden (Vangrieken et al., 2015). Im Folgenden wird in diesem Kontext der Begriff **Teamteaching** verwendet.

Die Mehrzahl an Studien beschäftigt sich mit der Zusammenarbeit von Lehrkräften und Sonderpädagog:innen (bspw. Jortveit & Kovač, 2022), insbesondere in inklusiven Settings. Andere Untersuchungen widmen sich der Zusammenarbeit zwischen zwei Lehrkräften, sowohl im schulischen (z.B. Decuyper et al., 2023) als auch im universitären Bereich (z.B. Álvarez et al., 2024). Viele Publikationen sind jedoch eher als Praxisberichte oder Fallbeispiele zu betrachten (bspw. Randall & Marangell, 2021; Rauterberg & Hauksdóttir, 2023). Befunde zur Zusammenarbeit zwischen Lehrkräften und Sozialarbeiter:innen liegen bisher nicht vor, lediglich der Abschlussbericht der ersten Förderperiode (Mitte et al., 2023) ermöglicht einen ersten Einblick in diese Form der Zusammenarbeit.

1.2.1 Internationaler Forschungsstand zur Wirksamkeit des Teamteachings

Im internationalen Bereich liegen vielfältige Untersuchungen zur Wirksamkeit von Teamteaching und verwandter Methoden vor. Zusammengefasst ist an dieser Stelle bereits festzuhalten, dass sich die **Befunde zwischen den Studien deutlich unterscheiden, so dass von einer generellen Effektivität nicht ausgegangen werden kann**. Zudem gestaltet sich die Studienlage insbesondere **hinsichtlich der methodologischen Qualität und der Standardisierung der implementierten Teamteaching-Konzepte sehr heterogen** (siehe dazu Weerdt et al., 2024). Große Unterschiede gibt es in der Auswahl der Outcome-Variablen, und viele Studien beschränken sich auf eine Wirkebene, bspw. die der Lehrkräfte (siehe dazu Vangrieken et al., 2015). Dadurch ist eine systematische Beurteilung im Rahmen von Meta-Analysen oder -Synthesen zusätzlich erschwert. So stellen Weerdt et al. (2024) fest, dass 61% der im Rahmen ihrer Meta-Synthese untersuchten Studien zum Teamteaching nur mittlere oder niedrige experimentelle Qualität aufweisen.

Laut einer Meta-Analyse von Vangrieken et al. (2015) scheinen **insbesondere die Lehrkräfte selbst zu profitieren**. Es konnten positive Effekte auf die Entwicklung der professionellen Identität von Lehrkräften (Álvarez et al., 2024; Tsybulsky & Muchnik-Rozanov, 2019) und auf die Beziehungsarbeit unter Lehrkräften (Rabin, 2020) gefunden werden. Vor allem bei einer voranschreitenden Dauer der Intervention fanden sich positive Effekte in der Routiniertheit der Zusammenarbeit (Decuyper et al., 2023), sowie in der professionellen Fähigkeit, die Bedürfnisse der Schüler:innen zu erkennen und auf sie einzugehen (Thompson & Schademan, 2019). Durch gegenseitiges Feedback lernen die beteiligten Lehrpersonen auch

voneinander (Jortveit & Kovač, 2022). Sichtbar wird dies auch durch den Erwerb und die Anwendung neuer Lehrkonzepte (Álvarez et al., 2024). Auf individueller Ebene der Lehrkräfte zeigte sich durch die Teamarbeit eine gesteigerte Selbstwirksamkeit (z.B. Bruce & Flynn, 2013; Mintzes et al., 2013) und Motivation (Vangrieken et al., 2015). Auch die Möglichkeit gegenseitigen Feedbacks wird als positiv aufgenommen (Álvarez et al., 2024). Die Beziehung zu den Schüler:innen veränderte sich in Richtung einer ressourcenorientierten und individuelleren Bewertung der Kinder (Hollins et al., 2004; Randall & Marangell, 2021). Es ergaben sich außerdem **positive Veränderungen in der gesamten Schulkultur wie eine verbesserte Kommunikation im gesamten Kollegium oder gesteigerte Offenheit** gegenüber alternativen Konzepten (Berry et al., 2005; Strahan, 2003; Vangrieken et al., 2015).

In Hinblick auf die Entwicklung der Schüler:innen ist die Forschungslage weniger deutlich. Es finden sich Untersuchungen mit **positiven Effekten** auf schulische Fertigkeiten wie Lese- oder Mathematikkompetenzen und allgemeine akademische Leistungen (Ebokaiwe et al., 2024; Vangrieken et al., 2015; Vembye et al., 2024) sowie auf die Lernmotivation der Schüler:innen (Anwar et al., 2021), aber auch auf soziale Fertigkeiten der Schüler:innen (Nevin et al., 2008; Rea et al., 2002). Es finden sich jedoch auch **Nullergebnisse** (Johnson et al., 1995), bspw. bei Jurkowski und Müller (2018), die über ein Schuljahr hinweg keine von Schüler:innen berichteten positiven Zugewinne in der Qualität der Lehre finden konnten, bis hin zu **negativen Effekten** (Zigmond & Baker, 1990). Auch in einer aktuellen Meta-Analyse von Stefanidis et al. (2023) ließen sich sowohl bei Schüler:innen mit als auch ohne Behinderung keine signifikanten Unterschiede in der akademischen Leistung gegenüber Klassen ohne Teamteaching finden. Im oben erwähnten Abschlussbericht der ersten Förderperiode (Mitte et al., 2023) ließen sich im Untersuchungszeitraum ebenfalls (noch) keine positiven Auswirkungen des STT auf die schulische Leistung oder die Fehltage der Schüler:innen beobachten. Positive Effekte fanden sich hier v.a. im emotional-sozialen Bereich, bspw. milderte höheres Vertrauen zur teamteachenden Person den Abfall schulischen Engagements über die Zeit ab (siehe dazu Mitte et al., 2023, S. 75–77).

1.2.2 Gelingensbedingungen

Verschiedene Review-Arbeiten (Murawski & Swanson, 2001; van Garderen et al., 2012) kommen zu dem Schluss, dass **zwar in der Mehrzahl von Studien positive Effekte zu finden seien, dass aber nicht generell davon ausgegangen werden kann, dass Teamteaching funktioniert**. Stattdessen sollten Gelingensbedingungen identifiziert werden, was allerdings bislang in nur wenigen Studien *systematisch* stattfand (Takala & Uusitalo-Malmivaara, 2012; Tanner-Smith & Wilson, 2013). Als potenzielle Moderator-Faktoren, welche die Wirksamkeit der Programme beeinflussen, werden einige konkrete Variablen als Ergebnis dieser Meta-Studien

diskutiert, so beispielsweise das Alter und Geschlecht der Schüler:innen. Zudem betrachten verschiedene Autor:innen innerhalb ihrer Projekte die Fragestellung der Gelingensbedingungen (Härkki et al., 2021; Jurkowski et al., 2020; Kokko et al., 2021; Rytivaara et al., 2019; Thompson & Schademan, 2019).

Hintergründe der Schüler:innen und Klassenzusammensetzung

Im Folgenden soll speziell noch einmal die Klassenzusammensetzung für ein potenziell gelingendes Teamteaching im Fokus stehen. Insbesondere im Forschungskontext zur Ätiologie von und Intervention bei deviantem Verhalten im Kindes- und Jugendalter liegt eine **Vielzahl an Studien vor, die den bedeutsamen Einfluss der Zusammensetzung von Gruppen untersuchte**. Dabei konnte gezeigt werden, dass Trainingsprogramme in Gruppen mit ausschließlich devianten Jugendlichen nicht nur im Mittel niedrigere Wirksamkeiten hatten, sondern dass sich in einer Vielzahl an Studien sogar negative Effekte zeigten, d.h. dass antisoziales Verhalten in diesen Gruppen anstieg trotz des Programms (McCord & Tremblay, 1992; Palinkas et al., 1996). Die Erklärungsansätze sind sowohl lerntheoretisch als auch sozialpsychologisch begründet. Sind Gruppen durch eine Vielzahl auffälliger Kinder und Jugendlicher definiert bzw. bestehen im Extrem nur aus diesen, so führt dies u.a. zu einer Isolierung von anderen Kindern und Jugendlichen (Dodge et al., 2006) und damit einem Fehlen von positiven Rollenmodellen und gesunden Beziehungsmustern. Es kann auch negative Labels und Selbstzuschreibungen und damit sich selbst-erfüllenden Prophezeiungen hervorrufen („Wenn ich zur Loserklasse gehöre, dann brauche ich mich auch nicht mehr anstrengen“).

Die gegenseitige Verstärkung negativen Verhaltens und einer Etablierung einer negativen Gruppennorm kann zu einem insgesamt negativen Klassenklima führen. Dishion und Tipsord (2011) führen ebenfalls die Bedeutsamkeit des sogenannten *peer contagion*, der sog. „Ansteckung“ durch Gleichaltrige, in sowohl natürlichen Interaktionen wie Freundschaften aber auch institutionalisierten Programmen und Trainings aus. Sie zeigen in ihrer Literaturübersicht, **dass nicht die Programme selbst ineffektiv waren, sondern der Kontext: der Einsatz ansonsten wirksamer Programme in homogenen Gruppen mit hohem auffälligem Verhalten kann nicht nur zu einer geringeren Effektivität führen, sondern sogar schädliche Effekte auf die Kinder und Jugendliche haben**. Homogene Gruppen stehen auch im Widerspruch zum Grundprinzip inklusiver Pädagogik, nämlich heterogene Gruppen gemeinsam zu beschulen (Lindner & Schwab, 2020).

Tandemzusammensetzung, Lehrkräfte und andere Fachkräfte

Bei Rexroat-Frazier und Chamberlin (2019) wurden die Gelingensbedingungen in Hinblick auf Teamteaching im Matheunterricht untersucht: vor allem die früh formulierten Zielsetzungen im Tandem, die optimierte Auswahl der/des Tandempartner:in sowie die Anwendung effizienter

Praktiken sowie spezifisch-mathematischer Lehrstrategien spielten dabei eine Rolle. Eine weitere kritische Variable in der Effektivität der Methode kann die Zusammenarbeit im Team sein.

Untersuchungen konnten eine Vielzahl an Gelingungs-Faktoren innerhalb des Tandems identifizieren (Bolam et al., 2005; Newmann & Wehlage, 1995; Takala & Uusitalo-Malmivaara, 2012; Vangrieken et al., 2015). Benannt wurden a) die gemeinsamen Wertvorstellungen und Visionen insbesondere hinsichtlich der (positiven) Ansichten über die individuellen Bedürfnisse und Potentiale ihrer Schüler:innen sowie b) positive Erwartungen hinsichtlich deren Kompetenzen, c) geteilte und unterstützende Verantwortung für die Fortschritte jeder/s Schüler:in inklusive Offenheit gegenüber neuen Konzepten. Auch die kritische Reflektion des eigenen professionellen Handelns (d) unter Nutzung dialogischer Verfahren wurde hervorgehoben, sowie e) die Verhandlungsfähigkeit bei interindividuell unterschiedlichen Herangehensweisen (Thompson & Schademan, 2019). Der Fokus auf gegenseitigem Lernen und Lernen voneinander (f), sowie g) die Anerkennung der Fähigkeiten und Kompetenzen der/des jeweils anderen (Jurkowski et al., 2023; Kokko et al., 2021) - die gegenseitige Stärkung und Unterstützung sind fundamental.

Einige Faktoren können zeitgleich förderlich und hinderlich sein, bspw. die Autonomie der beteiligten Personen (Vangrieken et al., 2015). Lehrkräfte als traditionelle Einzelkämpfende stünden vor einer besonderen Herausforderung, nämlich ihre bisherige Autorität zu teilen, ohne das Engagement zu verlieren. Die Freiwilligkeit der Maßnahme kann zudem maßgeblich Einfluss darauf haben, wie eng die Kooperation im Team ist: wenn das Teamteaching von oben „verordnet“ wird, kann sich das in geringerer Kooperation widerspiegeln (Vangrieken et al., 2015). Auch in der Studie von Moser Opitz et al. (2021) zeigte sich, dass die Lernunterstützung der Schüler:innen als höher wahrgenommen wird in der Kontrollgruppe ihres Experimentaldesigns, in der das Teamteaching weniger streng und nicht verpflichtend bzw. ohne standardisierte Vorgaben erfolgt, als in der Experimentalgruppe, in der die Ausführung des Teamteachings hinsichtlich bestimmter Details obligatorisch ist.

Deutsche Lehrkräfte standen in einer Untersuchung von Jurkowski et al. (2020) dem Konzept von Teamteaching generell sehr positiv gegenüber, und erkannten die zu aktivierenden Ebenen im Bildungssystem, an der jeweiligen Schule und in der Tandems-Dyade. Bezogen auf die Qualität im Tandem wurde zwar in einer Untersuchung von Vembye et al. (2024) kein signifikanter Effekt zwischen speziell ausgebildeten Co-Teacher:innen und nicht ausgebildeten Co-Teacher:innen gefunden, allerdings sprechen andere Befunde für die Wichtigkeit der Vorbereitung (Chitiyo, 2017), für die fachliche Passung der Tandempartner:innen und der Anwendung effektiver Praktiken (Rexroat-Frazier & Chamberlin, 2019).

Im Rahmen einer Befragung von Lehrkräften (Kokko et al., 2021) konnte herausgearbeitet werden, dass es wahrgenommene Unterschiede in den Chancen und Herausforderungen des Teamteachings gibt. Dabei wird zwischen Klassenlehrer:innen, Fachlehrer:innen und Sonderpädagog:innen unterschieden, da sie aufgrund von zeitlichen Rahmenbedingungen häufig unterschiedliche Beziehungen zur Klasse führen. Als erklärende Faktoren für die unterschiedlichen Ansichten der Lehrkräfte werden Variablen wie das Geschlecht (Kokko et al., 2021), das Alter (Stefanidis et al., 2019), die Co-Teaching-Erfahrungen der Lehrkräfte, die Häufigkeit des Co-Teachings pro Woche und die Aktivität als Klassenlehrer:in genannt (Kokko et al., 2021; Vangrieken et al., 2015).

Schlussfolgernd kann gesagt werden, dass die **Zusammenarbeit auf Augenhöhe** wohl den Kern des gelingenden Teamteachings darstellt (Kokko et al., 2021), damit das Mehr an Ressource optimal dafür genutzt werden kann, zu beobachten, was im Klassenraum geschieht. Je besser die Beziehung zueinander, desto positiver bewerten die Tandempartner:innen die Auswirkungen ihrer Zusammenarbeit (Stefanidis et al., 2019). In Bezug auf das STT könnten insbesondere die Wertschätzung der Zusammenarbeit, aber auch ihres individuellen Beitrags von Bedeutung sein, da bspw. die finanzielle Vergütung von Lehrkräften und Sozialarbeiter:innen stark voneinander abweichen und ggf. Konfliktpotenzial bieten könnte.

Schulebene

Auf der administrativen Ebene können Grundlagen für effektives Teamteaching geschaffen werden (Strogilos et al., 2023). Dabei sollte die gewählte Intervention und die Gestaltung der Zusammenarbeit sich immer am spezifischen Kontext der Schule, der Schüler:innen und der Lehrkräfte orientieren (Vangrieken et al., 2015). Ob und wann das Teamteaching im konkreten Fall jedoch eine geeignete Maßnahme sein kann, ist aufgrund der gemischten Befundlage nicht so leicht zu beantworten.

Gerade **im Kontext von Schulen in herausfordernden Lagen sind gezielte pädagogische Interventionen nützlich und wichtig**, da strukturelle und soziale Bedingungen besondere Herausforderungen im Schulalltag darstellen. In Schulen in herausfordernden Lagen lernen Kinder und Jugendlichen, denen das Lernen nicht leicht- oder sogar schwerfällt. Die Familienhintergründe sind häufig durch einen geringen formalen Bildungsstand, niedrigen sozioökonomischen Status und niedrige Erwerbstätigkeitsquoten charakterisiert (Duke et al., 2014). Und genauso fällt es auch den Schulen schwer, diese Kinder und Jugendlichen erfolgreich zu beschulen. Das liegt oft an dem Umfeld der Schule, welches nicht genug Anregung oder die Möglichkeit für Kooperationen bietet, meist ist die Infrastruktur schlecht (Böttcher et al., 2022). Da auch die Schulen, die am STT teilnahmen und teilnehmen, als Schulen in herausfordernden Lagen zu bezeichnen sind, soll das Augenmerk darauf gelegt werden.

Sowohl in quantitativen als auch in qualitativen Studien und Reviews sticht die **Ressource Zeit** hervor, um das gemeinsame Unterrichten einerseits im Team und andererseits individuell vor- und nachzubereiten (bspw. Jurkowski et al., 2023; Stefanidis et al., 2019; Strogilos et al., 2023; Takala & Uusitalo-Malmivaara, 2012). Auch die Planbarkeit bezüglich einer längerfristigen Zusammenarbeit beeinflusst die Zusammenarbeit im Tandem (Bolam et al., 2005). In Bezug auf die Implementation spielt neben der oben erwähnten Freiwilligkeit der Teilnahme (Vangrieken et al., 2015) die Ausstattung der Tandems eine wichtige Rolle (Jurkowski et al., 2023; Stefanidis et al., 2019). Insbesondere bei inter-disziplinären Teamteaching-Formen kann die multi-professionelle Zusammenarbeit innerhalb der Schule, aber auch die Unterstützung durch außerschulische Expert:innen förderlich sein (Jurkowski et al., 2023).

1.3 Methodisches Vorgehen

Die Evaluation umfasst Daten aus eigenen quantitativen Erhebungen auf der Ebene der Schüler:innen (Fragebögen und Übersichten der Noten und Fehltage, siehe **TEIL II – SCHÜLER:INNEN-BEFRAGUNG** und **TEIL III – ÜBERSICHTEN DER NOTEN UND FEHLTAGE**) und Klassen (Fragebögen von Tandemlehrkräften und Teamteacher:innen: **TEIL IV – BEFRAGUNG DER TEAMTEACHENDEN UND KLASSENLEHRKRÄFTE**). Die Ergebnisse dieser Erhebungen werden im Folgenden berichtet. Wo möglich, wurden Daten kombiniert und gemeinsam analysiert.

1.3.1 Auswahl der Interventions- und Kontrollgruppen

Aufbauend auf den Erkenntnissen der ersten Projektperiode wurde ein Kontrollgruppendesign angestrebt. Dafür wurden an allen am STT beteiligten Schulen abgefragt, welche Klassen aktuell am STT beteiligt sind. Angestrebt wurde eine Größe von etwa 20-25 Klassen in jeder Erhebungsgruppe. Wir weisen darauf hin, dass es sich hier um eine Feldstudie handelt und nicht um ein vollständig randomisiertes Design, wie sonst in Evaluationsstudien üblich (Gollwitzer & Jäger, 2014). Die Auswahl der Klassen erfolgte so, dass für jede an der Erhebung beteiligte STT-Klasse eine „Kontrollgruppe“ gleichen Alters und ungefähr gleicher Klassengröße vorhanden war. Die Auswahl der STT-Klassen erfolgte basierend auf Informationen der KSB zu aktuell existierenden STT-Klassen sowie in Rücksprache mit den beteiligten Schulen. Im Rahmen von Telefonaten mit den Schulleitungen und Teamteachenden wurden die Informationen der KSB mit den Schulen abgeglichen und Rückfragen zu den u.g. Auswahlkriterien gestellt. Die Kontrollgruppen wurden nachfolgend ausgewählt mit dem Ziel, an jeder Schule sowohl STT- als auch Kontrollgruppen an der Erhebung zu beteiligen, um auch eine Vergleichbarkeit hinsichtlich der sozialen Lage der Schulen zu gewährleisten. Im Verlauf der Evaluation kam es jedoch zu, dass einige Schulen nur mit STT- oder nur mit Kontrollgruppen beteiligt waren (siehe dazu den folgenden Abschnitt 1.3.2).

Als **Interventionsklassen** wurden diejenigen Klassen ausgewählt, die a) nicht bereits an den Erhebungen der ersten Förderperiode teilgenommen hatten, und laut Aussage der beteiligten Schulen b) zu Beginn des Schuljahres 22/23 mit dem STT begonnen hatten, c) höchstens der 7. Jahrgangsstufe angehörten, d) nicht an jahrgangsübergreifendem Unterricht beteiligt waren sowie e) auch zu den zwei darauffolgenden Erhebungszeitpunkten STT-Klassen bleiben sollten.³ Als **Kontrollgruppen** kamen alle Klassen an den Projektschulen in Frage, die a) nicht zu einem früheren Zeitpunkt schon STT-Klassen gewesen waren und laut Aussage der beteiligten Schulen b) nicht an jahrgangsübergreifendem Unterricht beteiligt waren und c) nicht zu den zwei darauffolgenden Erhebungszeitpunkten STT erhalten sollten.

Wie sich die Anzahl der Klassen in der Interventions- und Kontrollgruppe über die Erhebung hinweg veränderte, ist in **Tabelle 1** dargestellt. Welche Konsequenzen diese Veränderungen für die Auswertung der Daten hatte, wird im folgenden Abschnitt dargestellt. Im Folgenden nutzen wir die Unterscheidung von STT (Interventionsgruppe: Klassen mit STT) versus KG (Kontrollgruppe: Klasse ohne STT).

1.3.2 Aufteilung in *intention to treat* und *per protocol*

Über den Erhebungszeitraum hinweg veränderte sich die Verteilung der Klassen zwischen der STT- und Kontrollgruppe mehrfach. Wie in **Tabelle 1** zu sehen ist, kam es mehrfach zu Beendigungen der Maßnahme, entweder weil Teamteachenden-Stellen vakant wurden oder weil die Teamteachenden in andere Schulklassen derselben Schulen wechselten. Auch einige ursprünglich der Kontrollbedingung zugeteilten Klassen erhielten nach T1 STT.

Tabelle 1 Klassen mit und ohne STT im Laufe der Evaluation

	Mit STT	Ohne STT
T1	20	30
T2	22	28
T3	18	32

Um die für Längsschnittanalysen erforderliche Stichprobengröße nicht weiter zu reduzieren, wurden Klassen mit Statuswechsel jedoch nicht von der Erhebung ausgeschlossen. Stattdessen wurde eine Unterteilung der beteiligten Klassen nach *intention to treat* (ITT) und *per protocol* (PP) vorgenommen (Erklärung siehe unten). In Evaluationsstudien üblich, ermöglicht diese Unterteilung zwei unterschiedliche Ansätze zur Auswertung der Daten. Wie viele Klassen zu

³ Grund dafür war, dass viele Klassen der 10. Jahrgangsstufe ebenfalls vom ESF geförderte Maßnahmen zur Berufsberatung o.ä. erhalten und nicht zeitgleich zwei Maßnahmen in einer Klasse gefördert werden – weswegen in vielen Klassen das STT mit Ende der 9. Klasse auch beendet wird.

den jeweiligen Erhebungszeitpunkten zur STT- oder Kontrollgruppe gehörten und wie viele Klassen in die ITT- und PP-Bedingung aufgenommen wurden, ist **Tabelle 2** zu entnehmen.

Tabelle 2 Aufteilung der Klassen in STT- und Kontrollgruppe

	STT-Gruppe	Kontrollgruppe
Vor T1 geplante Auswahl (ITT)	25	25
PP	14	21

Beim **ITT-Ansatz** werden alle Teilnehmenden in die Analyse einbezogen, so wie sie ursprünglich den Untersuchungsgruppen zugeteilt wurden – unabhängig davon, ob sie die Intervention vollständig erhalten oder vorzeitig abgebrochen haben. Dieser Ansatz spiegelt die Wirksamkeit einer Maßnahme unter realen Bedingungen wider und vermeidet Verzerrungen durch selektiven Ausfall.

Der **PP-Ansatz** berücksichtigt dagegen nur jene Teilnehmenden, die die Evaluation gemäß den Vorgaben vollständig durchlaufen haben. Wie in anderen Wirksamkeitsstudien mit länger andauernden Interventionen üblich, haben wir dabei keine 100%ige Implementation vorausgesetzt, sondern eine Mindestdosis erhaltener Intervention als Bedingung gesetzt (Kipping et al., 2014; Leahy et al., 2019). In der vorliegenden Stichprobe bedeutet dies, dass nur die Klassen als STT-Klassen in der Analyse berücksichtigt werden, die über die gesamte Beobachtungszeit hinweg in jedem Messzeitraum/Schuljahr mindestens 91 Tage eine:n Teamteacher:in hatten. Als Kontrollgruppe wurden wiederum nur die Klassen berücksichtigt, die durchgängig kein oder weniger als 91 Tage STT hatten. Dadurch kann genauer untersucht werden, wie wirksam die Intervention unter idealen, optimalen Bedingungen ist. Beide Ansätze ergänzen sich, da sie unterschiedliche Perspektiven auf die Wirksamkeit und Umsetzbarkeit des STT bieten.

1.4 Genehmigung der Erhebungen

Alle Erhebungsmaterialien wurden in enger Zusammenarbeit mit der KSB und dem Thüringer Ministerium für Bildung, Wissenschaft und Kultur (TMBWK) entwickelt und an die individuellen Bedarfe der Zielgruppen angepasst. Diese wurden inklusive einer detaillierten Beschreibung des Erhebungsablaufs sowie Änderungen am Ablauf oder der Erhebungsmaterialien im Vorfeld jeder Erhebungswelle beim Schulamt angemeldet, welches am 26.04.2023 nach §57 Abs. 5 ThürSchulG eine Genehmigung für die „Evaluation ‚Sozialpädagogisches Teamteaching‘“ erteilte.

1.5 Forschungsdatenmanagement

Die Erhebung und Auswertung von Daten auf Schüler:innen-Ebene ist im wissenschaftlichen Bereich zurecht nur mit vorheriger Genehmigung durch die Schulämter möglich und setzt einen **sensiblen Umgang bei der Erhebung, Speicherung und Auswertung** voraus. Zusätzlich zu den im Projekt vorhandenen Kompetenzen wurde daher eine Kooperation mit dem EAH-eigenen „Team Forschungsdatenmanagement“⁴ eingegangen.

Insbesondere bei der Erhebung der Übersichten der Noten und Fehltage (ÜNuF) wurden verschiedene Strategien entwickelt, um eine Erhebung, Speicherung und Auswertung der Daten auf **höchstem datenschutzrechtlichem Niveau** zu gewährleisten (siehe dazu auch Abschnitt 3.1). Dabei wurde zusätzlich darauf geachtet, den Aufwand der Dateneingabe auf Seiten der Lehrkräfte und Teamteachenden so gering wie möglich zu halten und sinnvoll an den Schuljahresablauf anzupassen.

1.6 Inklusion, Diversität und Nachhaltigkeit

Die gesamte Evaluation, insbesondere aber die Befragung der Schüler:innen, fand unter besonderer Berücksichtigung **alters-, geschlechts-, migrations- und diskriminierungssensibler Maßnahmen** sowie unter **Einhaltung geltender Datenschutzrichtlinien** statt. Regelmäßig wurde die Einhaltung der im Projektantrag aufgeführten Querschnittsziele im Projektteam reflektiert und die Forschungsarbeit und -kommunikation dahingehend angepasst. Die Einhaltung der Querschnittsziele „**Gleichstellung der Geschlechter**“ sowie „**Chancengleichheit und Nichtdiskriminierung**“ ist für uns selbstverständlich. So achten wir auf gendergerechte und diskriminierungsfreie, aber auch leicht verständliche Sprache. Auch für die hiermit beantragte Maßnahme knüpfen wir an die Schaffung diskriminierungssensibler Erhebungsbedingungen an. Als Psycholog:innen und Soziolog:innen sind wir uns der Stereotype bezüglich Geschlecht und weiterer sozialer Kategorien und ihrer Wirkung sehr bewusst, weshalb wir diese aktiv in unserer Lehre und Forschung thematisieren.

Die Elternbriefe wurden in **leichter Sprache**, sowie auf Englisch, Arabisch und Russisch ausgegeben.⁵ Die Schüler:innen-Fragebögen waren auf Deutsch und Russisch⁶ verfügbar. Der Wortlaut einzelner Items wurde gegenüber den Originalversionen vereinfacht oder an den Sprachgebrauch der Schüler:innen angepasst. Bezogen auf diversitätssensible Aspekte erfassen wir z.B. die Migrationsgeschichte von Schüler:innen anhand mehrerer Maße, bspw. darüber,

⁴ Über das Projekt: <https://www.eah-jena.de/forschung/fdm>

⁵ Eine stichprobenartige Umfrage in den teilnehmenden Klassen hatte ergeben, dass die meisten der DaZ-Schüler:innen aus arabisch- und ukrainisch-/russisch-sprachigen Familien stammen.

⁶ Laut den beteiligten Schulen wurden insbesondere für Schüler:innen aus der Ukraine Übersetzungen benötigt, wobei Russisch von vielen ukrainischen Familien gesprochen und auch für ukrainisch-sprachige Schüler:innen leicht verständlich ist.

welche Sprache(n) in den Familien vorrangig gesprochen werden (siehe dazu 2.2.2). So erfassen wir zum einen die Diversität der individuellen Hintergründe und ermöglichen andererseits den Schüler:innen, abseits von ungenauen Kategorien wie „Migrationshintergrund“ mehrere Aspekte ihrer Identität abzubilden. Am Ende des Fragebogens (FB) wurde abgefragt, wie gut die Schüler:innen den Fragebogen verstanden hatten.

Aus Gründen der besseren Lesbarkeit wurde in den Items auf Gendern verzichtet und ausschließlich die männliche Form verwendet (z.B. „deine Lehrer“). Am Beginn des Fragebogens fand sich eine Erklärung, dass mit dieser Form **alle Geschlechter** gemeint sind. Bei der Angabe des Geschlechts hatten die Schüler:innen die Möglichkeit, aus mehreren Optionen (Junge, Mädchen, trans, nicht-binär) auszuwählen oder in einem Freifeld selbst ihr Geschlecht anzugeben. Während der Erhebung wurde auf die Aufteilung von Schüler:innen nach Geschlecht verzichtet.

Die vorliegende Schüler:innen-Stichprobe umfasste eine **große Altersspanne**, daher wurden alle Materialien und Abläufe so gestaltet, dass sie sowohl für jüngere als auch für ältere Schüler:innen zugänglich waren. Die Fragebögen wurden eher kindgerecht gestaltet. So wurden bspw. die Antwortskalen durch Smileys untermalt, die bereits in anderen Erhebungen zum Einsatz gekommen waren (Hall et al., 2016). Alle Skalen wurden einleitend kurz erläutert. Wurde während einer Erhebung deutlich, dass einzelne Items schwer verständlich waren, wurden diese Items bzw. die Erläuterung zu darauffolgenden Erhebungen sprachlich angepasst. Bevor die Schüler:innen mit dem Ausfüllen der Fragebögen begannen, wurde der Ablauf altersgerecht und anhand eines Beispiel-Items erklärt.

Die Schüler:innen hatten jederzeit die Möglichkeit, Fragen zu stellen, die Teilnahme zu pausieren oder abubrechen. In einigen wenigen Fällen wurden in diesem Fall das STT-Tandem oder andere Lehrkräfte hinzugezogen. Die Beantwortungsdauer für die Fragebögen wurde großzügig geplant, damit auch Schüler:innen mit **erhöhtem Hilfsbedarf** beim Lesen ausreichend Zeit zum Ausfüllen der Fragebögen hatten. Schulbegleiter:innen durften während der Erhebung anwesend sein und bei Bedarf den von ihnen betreuten Schüler:innen assistieren, jedoch nicht bei inhaltlichen Fragen, diese wurden durch das Erhebungsteam geklärt. Waren Schüler:innen eher fertig, erhielten sie vom Erhebungsteam Mandalas zum Ausmalen oder konnten die Zeit nutzen, um Hausaufgaben zu machen. Allgemein wurde darauf geachtet, dass die Lautstärke im Klassenzimmer nicht zu laut wurde, in Einzelfällen wurden Schüler:innen auseinander gesetzt.

Während der Erhebung achteten die Erhebungsleiter:innen stets darauf, dass die **Privatsphäre** der Schüler:innen beim Ausfüllen der Fragebögen gewährleistet ist. Aufkommende Fragen wurden durch die Erhebungsleiter:innen vertraulich beantwortet. Die Gestaltung der

Erhebungsmaterialien sowie die Eingabe der Daten wurde so gestaltet, dass keine Rückschlüsse mehr auf Name, Klasse oder Schule der Schüler:innen möglich wurden.

Das Querschnittsziel „**Ökologische Nachhaltigkeit**“ wurde insbesondere über einen fast vollständigen Wechsel von einer Erhebung mit ausgedruckten Fragebögen (siehe dazu Mitte et al., 2023) hin zu einer Erhebung mittels Online-Fragebogen erfüllt.⁷ Dafür wurde auf die an den Schulen vorhandene Infrastruktur (v.a. Tablets) zurückgegriffen.

⁷ An einer Schule war keine entsprechende Infrastruktur vorhanden, sodass diese komplett mit Papierfragebögen erhoben wurde. Zudem wurde in Einzelfällen (Probleme mit dem Internet o.ä.) auf Papier-Fragebögen zurückgegriffen.

TEIL II – SCHÜLER:INNEN-BEFragung

2.1 Durchführung und Ablauf der Erhebung

Das Forschungsdesign zur Schüler:innenbefragung umfasst quantitative Daten aus Erhebungen mit Klassen mit und ohne STT. Wie bereits im Abschnitt 1.3 erläutert, wurde das ursprünglich geplante Erhebungsdesign erheblich durch Interventionsabbrüche oder -wechsel zwischen den Klassen eingeschränkt. Um dennoch Aussagen aus dem Vergleich zwischen Klassen mit und ohne STT sowie über mögliche Einflüsse durch die Interventionsdauer treffen zu können, wurden separate Analysen für die Aufteilung nach ITT und PP durchgeführt.

2.1.1 Durchführung der Erhebung

Waren die Erhebungen der ersten Förderperiode noch stark durch die pandemiebedingten Schulschließungen und Kontaktbeschränkungen geprägt (siehe Mitte et al., 2023), konnten alle Erhebungen wie geplant vor Ort in den Schulen stattfinden. Zwischen dem 11.05. und 22.06.2023 wurde die erste Erhebung (T1) durchgeführt, die zweite Erhebung (T2) fand zwischen 04.03. und 17.04.2024 statt und die dritte Erhebung (T3) zwischen dem 24.02. und 28.03.2025.

Im Vorfeld von T1 wurden an die Schulen **Einverständniserklärungen** für Schüler:innen bzw. deren Erziehungsberechtigte geschickt, die durch die Klassenleitungen und/oder Teamteacher:innen verteilt und wieder eingesammelt wurden.⁸ Mit dem Einverständnis der Erziehungsberechtigten wurden Schüler:innen für die Erhebung eingeplant. Die Erhebungen wurden von wissenschaftlichen Mitarbeiter:innen und geschulten studentischen Hilfskräften durchgeführt, Auffälligkeiten und Probleme während der Erhebung wurden protokolliert, im Team und ggf. mit der Klassenleitung der jeweiligen Klasse nachbesprochen. Die Lehrkräfte/Teamteacher:innen verließen gemeinsam mit Schüler:innen, die keine Einverständniserklärung hatten, zu Beginn der Erhebung den Raum und kehrten erst danach wieder für eine Übergabe in die Schulklasse zurück.

Die Schüler:innen wurden zusätzlich zur Einverständniserklärung altersgerecht über Ziel, Ablauf und Datenschutz in der Befragung informiert und es wurde erläutert, dass sie zu jedem Zeitpunkt (auch nach der Befragung) unabhängig von der Entscheidung ihrer Erziehungsberechtigten von ihrer Teilnahme zurücktreten können. Während der gesamten Erhebung wurde darauf geachtet, dass Schüler:innen in Ruhe und für sich die Fragen beantworten konnten. Inklusionsassistent:innen von einzelnen Schüler:innen durften während der Erhebung anwesend sein, wurden jedoch angewiesen, nur auf Nachfrage der ihnen

⁸ Zu T2 wurde dies für neu hinzugekommene Schüler:innen wiederholt, zu T3 jedoch nicht mehr.

zugeordneten Schüler:innen zu assistieren und auch sonst nicht in den Erhebungsablauf einzugreifen.

Im Online-Fragebogen waren mehrere Filterfragen enthalten, sodass bspw. Fragen zum STT nur denjenigen Schüler:innen angezeigt wurden, die auch angaben, aktuell STT in der Klasse zu haben. Die Erhebung in den Schulklassen der teilnehmenden Schulen fand im Rahmen eines regulären Schultages statt und dauerte pro Klasse maximal zwei Schulstunden, bei dem Fragebogen für die Kontrollgruppe wurden aufgrund des kürzeren Fragebogens teilweise eine Schulstunde eingeplant. Bevor die Klasse wieder an die Klassenleitung übergeben wurde, erhielten alle Schüler:innen kleine Dankeschön-Geschenke (Schreibutensilien) für ihre Teilnahme.

Alle an einer Projektschule teilnehmenden Schulklassen wurden im Laufe eines Schultages erhoben. Die Dauer der Erhebungszeiträume ist Abschnitt 1.3 zu entnehmen. Es wurde angestrebt, dass die Dauer zwischen den Erhebungszeitpunkten für alle teilnehmenden Schulen in etwa gleich war. Leichte Schwankungen in der Länge dieser Zeiträume können jedoch in den gewählten Multilevel-Analysen berücksichtigt werden und ermöglichen unabhängig davon eine Schätzung der individuellen Verläufe.

2.1.2 Stichprobenbeschreibung

In

Tabelle 3 sind die demographischen Kennwerte der Gesamtstichprobe dargestellt. In der zweiten Spalte ist angegeben, von wie vielen Schüler:innen zu der betreffenden Variable Angaben vorliegen. Welcher Messzeitpunkt (MZP) dabei priorisiert wurde (z.B. T1 bei Alter), ist den Anmerkungen unter der Tabelle zu entnehmen. Welche Klassenstufen beteiligt waren, ist ebenfalls **Tabelle 4** zu entnehmen.

In den Angaben spiegelt sich die Diversität der Gesamtstichprobe hinsichtlich der familiären und kulturellen Hintergründe der teilnehmenden Schüler:innen wider. Von den 23.8% der Schüler:innen, die zu T3 angaben, in einem anderen Land als Deutschland geboren zu sein, wurden 40 verschiedene Geburtsländer genannt. Am häufigsten waren darunter Kinder aus Syrien (7.4%) und der Ukraine (2.9%) vertreten. Auch bei den Angaben zum Geburtsland der Eltern war der Anteil von Schüler:innen mit nicht in Deutschland geborenen Elternteilen recht hoch (Elternteil 1/Mutter: 25.2%, Elternteil 2/Vater: 28.2%), wobei hier auch viele Kinder keine Angaben machen konnten oder wollten.

Wie unterschiedlich die Wohnsituation bzw. das Wohnumfeld der teilnehmenden Schüler:innen ist, spiegelt sich auch darin wider, dass alle der von uns aufgeführten Optionen mindestens ein Mal ausgewählt wurden. Auffällig ist, dass mehr als ein Viertel der Schüler:innen

ausschließlich bei ihrer Mutter (25.1%) wohnen. Der Großteil der Schüler:innen gab an, nur wenige Bücher zu besitzen (26.2%), als maximaler Bildungsabschluss der Eltern wurde am häufigsten der Realschulabschluss genannt (49.5%).

Tabelle 3 Beschreibung der Schüler:innen-Stichprobe

Variable	Ausprägungen / Kennwerte	absH	relH
Gesamtzahl Schüler:innen	Teilnahme zu T1	720	
	Teilnahme zu T2	703	
	Teilnahme zu T3	650	
	Teilnahme zu allen MZP	468	
Alter zu T1	Altersspanne	10-16	
	<i>M (SD)</i>	12.1 (1.1)	
Geschlecht ^a	<i>n</i>	858	
	männlich	450	52.4
	weiblich	390	45.5
	nicht-binär	6	.7
	trans	9	1.0
	Andere	3	.3
Geburtsland ^a	<i>n</i>	866	
	Deutschland	656	75.8
	anderes Land	210	24.2
Geburtsland Elternteil 1 (Mutter) ^{a, b}	<i>n</i>	775	
	Deutschland	553	71.4
	anderes Land	222	28.6
Geburtsland Elternteil 2 (Vater) ^{a, b}	<i>n</i>	767	
	Deutschland	518	67.5
	anderes Land	249	32.5
Anzahl der Geschwister zu T3	<i>n</i>	814	
	keine Geschwister	100	12.3
	1 Geschwisterkind	225	27.6
	2 Geschwister	198	24.3
	3 Geschwister	118	14.5
	mehr als 3 Geschwister	173	21.3
Maximaler Bildungsabschluss der Eltern ^a	<i>n</i>	666	
	kein Schulabschluss	31	4.7
	Grundschule	9	1.4
	Hauptschulabschluss	82	12.3
	Realschulabschluss	330	49.5
	Abitur	169	25.4
	Universitätsabschluss	45	6.8
Wohnsituation: „Ich wohne...“ ^a	<i>n</i>	812	
	zusammen mit meinen Eltern	494	60.8
	nur mit meiner Mutter zusammen	221	27.2
	nur mit meinem Vater zusammen	28	3.4
	bei anderen Verwandten	11	1.4
	zusammen mit Freunden	1	.1
	allein	1	.1
	mit anderen Personen	22	2.7

	abwechselnd bei meiner Mutter und meinem Vater	32	3.9
	bei meinen Pflegeeltern	2	.2
Bücherfrage ^{a, c}	<i>M (SD)</i>	2.3 (1.4)	

Anmerkungen. absH = absolute Häufigkeit, relH = relative Häufigkeit

^a Wenn Angaben zwischen den einzelnen MZP voneinander abwichen oder Daten fehlten, wurde die Angabe zu T3 priorisiert. ^b Wenn Kinder Angaben zu gleichgeschlechtlichen Elternpaaren machen wollten, konnten sie dieselben Antwortfelder nutzen. ^c 6 Antwortoptionen: 0 = Keine Bücher, 1 = Sehr wenige (1-10), 2 = genug, um ein Regalbrett zu füllen (11-25), 3 = genug, um ein Regal zu füllen, 4 = genug, um 2 Regale zu füllen (101-200), 5 = sehr viele Bücher (über 200).

2.1.3 Rücklauf- und Teilnahmequoten

Mithilfe der beteiligten Schulleitungen, Teamteachenden und Lehrkräften konnte über alle Messzeitpunkte hinweg ein hoher Anteil der Schüler:innen für die Erhebung gewonnen werden. Die durchschnittlichen Werte für die Klassengrößen, Rücklaufquoten (Anteil von Schüler:innen mit vorliegender Einverständniserklärung) und Anwesenheitsquoten (Anteil der zur Erhebung anwesenden und teilnehmenden Schüler:innen mit Einverständniserklärung) sind im unteren Teil der **Tabelle 4** zu sehen. Beide Anteile beziehen sich auf die Klassenstärke zum jeweiligen Zeitpunkt. Zudem ist getrennt nach STT- und PP-Bedingung angegeben, welche Klassenstufen (zu T1) in der Erhebung vertreten waren.

Wie in Evaluationsstudien mit Kontrollgruppendesign üblich, lag die Rücklaufquote in der Kontrollgruppe etwas niedriger als in den Interventionsklassen (siehe dazu bspw. Gollwitzer & Jäger, 2014). Von Schüler:innen, die zwischen T1 und T2 neu in der Erhebungsklasse kamen, wurden Einverständniserklärungen nachgefordert. Schüler:innen, die nach T2 hinzukamen, wurden nicht mehr in die Erhebung aufgenommen – sichtbar an den niedrigeren Rücklaufquoten zu T3. Eine Beteiligung zu einem einzigen Zeitpunkt wäre für das längsschnittliche Design nicht sinnvoll gewesen. Die höchste Rücklaufquote waren 80.2% in der STT-Gruppe in der ITT-Bedingung zu T2, die niedrigste 59.0% in der Kontrollgruppe der ITT-Bedingung zu T3.

Tabelle 4 Durchschnittliche Klassengrößen, Rücklauf- und Anwesenheitsquoten

			ITT		PP	
			KG	STT	KG	STT
T1	Klassenstufe	5	11	11	8	6
		6	9	9	8	7
		7	5	5	5	1
T2	Klassengröße		21.7	22.8	22.0	23.3
	Rücklaufquote		64.5%	75.8%	63.6%	76.1%
	Anwesenheitsquote		61.3%	72.9%	60.0%	68.9%
	Klassengröße		22.2	22.8	22.3	23.2
	Rücklaufquote		69.6%	80.2%	70.1%	79.4%

T3	Anwesenheitsquote	57.9%	67.4%	57.4%	66.0%
	Klassengröße	23.2	23.0	23.2	23.4
	Rücklaufquote	59.0%	70.7%	58.7%	69.7%
	Anwesenheitsquote	52.1%	60.6%	51.1%	61.1%

2.2 Entwicklung der Fragebögen

Die Analyse der Schüler:innen-Fragebögen bildet das Herzstück der Evaluation des STT. Die enthaltenen Skalen und Items geben Aufschluss über die **individuellen Hintergründe der Schüler:innen, sowie über deren subjektives Erleben des Schulalltags allgemein** und speziell in der STT-Gruppe über **das Erleben des STT** und dessen Auswirkungen auf individueller und auf Klassenebene.

Die Auswahl der Erhebungsinstrumente basierte stark auf den Erfahrungen und Ergebnissen der ersten Förderperiode, ein weiterer Schwerpunkt lag auf Skalen, die einen Vergleich zwischen Kontroll- und STT-Gruppe ermöglichen. Theoretische Grundlage für Interpretation und Einordnung der Ergebnisse bildet die *Self-Determination Theory* (Selbstbestimmungstheorie, Ryan & Deci, 2000), welche im folgenden Abschnitt kurz erläutert wird.

2.2.1 Theoretischer Hintergrund der Schüler:innen Fragebögen

Ausgangspunkt der Evaluation bildet die Frage, wie Schüler:innen „erfolgreich“ am Unterricht teilhaben können und welche Ressourcen ihnen dafür im schulischen Umfeld zur Verfügung gestellt werden können. Dabei sind insbesondere die Ressourcen von Interesse, die geeignet sind, strukturelle Benachteiligungen auszugleichen und so allen Schüler:innen eine gleichberechtigte, aktive Teilhabe am Unterricht ermöglichen.

Neben Erkenntnissen aus der internationalen Forschung zu Teamteaching (siehe Abschnitt 1.2) orientiert sich die Auswahl der Skalen an der *Self-Determination Theory* (SDT) nach Ryan und Deci (2000). Die SDT ist eine der am besten wissenschaftlich belegten Theorien der Sozial- und Bildungspsychologie. Kerngedanke dieser Theorie ist es, dass akademische Leistung und psychologisches Wohlergehen im Schulkontext voraussetzen, dass die selbstbestimmte (oder: **intrinsische**) Motivation der Schüler:innen gefördert wird. Je besser die Motivation der Schüler:innen, desto höher ist ihre Lernfreude, und desto besser die Qualität ihrer Teilnahme bzw. Mitarbeit am Schulleben (nachfolgend: **Schul-Engagement**). Dies wirkt sich schließlich auch positiv auf ihre akademische Leistung aus (Nouwen & Clycq, 2020). Schulabbruch hingegen ist nach dieser Theorie der natürliche letzte Schritt eines **Abwendens von der Schule** aufgrund mangelnder intrinsischer Motivation, zunehmend schlechterem Schul-Engagement, gekoppelt

mit einem letztendlichen Wegbrechen der noch verbliebenen extrinsischen Motivation (wie z.B. Schulzwang, elterlicher Druck).

Die wichtigste **Forschungsfrage**, die sich aus der SDT ergibt, ist dementsprechend die Frage danach, **wie Schulen und Lehrkräfte die Motivation Ihrer Schüler:innen fördern und erhalten können**. Hier postuliert die SDT **drei Grundbedürfnisse**, deren Erfüllung Voraussetzung für die adaptive Motivationsentwicklung ist: **Autonomie, Kompetenzerleben und soziale Eingebundenheit** (Ryan & Deci, 2000). Im Schulkontext versteht man unter sozialer Eingebundenheit unter anderem ein Gefühl der Zusammengehörigkeit und Kooperation innerhalb der Klasse, sowie zwischen Schüler:innen und Lehrern. Kompetenzerleben bezieht sich auf das subjektive Gefühl der Schüler:innen, ihre Aufgaben bewältigen zu können und Erfolge zu haben. Autonomie ist per Definition in der Schule oft eingeschränkt, aber zeigt sich zum Beispiel in dem Gefühl, „man selbst“ sein zu können, Auswahlmöglichkeiten zu bekommen oder individuelle Lernmethoden verwenden zu dürfen.

Diese Theorie ist durch jahrzehntelange Forschung empirisch belegt worden, und stellt daher ein ideales Grundgerüst für die Evaluation des STT dar. Es wird angenommen, dass das STT **Auswirkung auf die Person selbst und somit die zentralen Variablen der SDT** hat (v.A. Bedürfnisse, intrinsische und extrinsische Motivation, Schul-Engagement), und somit werden **langfristige positive Entwicklungen in der Schulleistung und dem Verhalten der Schüler:innen, sowie letztendlich eine Verringerung von Schulabbruch** erwartet. Im Vergleich zur Kontrollgruppe sollte sich hingegen zeigen, dass sich hier ein signifikanter Vorteil von Schüler:innen mit STT gegenüber Schüler:innen ohne STT zeigt.

2.2.2 Verwendete Skalen

Die Fragebögen für die Schüler:innen setzen sich zusammen aus selbst entwickelten Items und standardisierten, validierten psychologischen Skalen, die teilweise leicht an die Zielgruppe bzw. Fragestellung angepasst wurden (siehe dazu Abschnitt 2.1.1). Die Auswahl der zu erhebenden Konstrukte stützt sich auf internationale und nationale Befunde aus den Forschungsbereichen Schulen in herausfordernden Lagen (bspw. Holtappels et al., 2017), zu Dropout und Schulabstinz (bspw. Gubbels et al., 2019; Hennemann et al., 2010) sowie STT (bzw. verwandte Teamteaching-Konzepte; siehe dazu Solis et al., 2012). Zusätzlich orientierten sich die Autor:innen an dem Manual zum STT (Kindersprachbrücke Jena e. V., 2019)⁹ und an den Erkenntnissen der vorangegangenen Förderperiode (Mitte et al., 2023). Theoretischer Hintergrund und Anpassung der verwendeten Skalen werden in den folgenden Abschnitten dargestellt. Wenn nicht anders beschrieben, wurden Skalen zu allen Messzeitpunkten unverändert fortgeführt. Im

⁹ Das 2025 erschienene überarbeitete Manual der Kindersprachbrücke konnte bei der Skalenauswahl noch nicht berücksichtigt werden.

Anhang in **Tabelle 20** sind für alle verwendeten Skalen Mittelwerte, Standardabweichungen und McDonald's Omega (ω) als Maß der internen Konsistenz angegeben. Dabei wird ein Wert über .70 als akzeptabel, ein Wert von .80 als gut und ein Wert von .90 oder größer als ausgezeichnet betrachtet. Die hier verwendeten Skalen erzielten in der Gesamtstichprobe überwiegend akzeptable bis ausgezeichnete Omega-Werte.

Skalen zum STT und zu Lehrpersonen

Ein häufiges Thema in den Schüler:innen-Interviews der ersten Förderperiode (siehe dazu Mitte et al., 2023, S. 34) war, dass einige Schüler:innen über ein sehr starkes **Vertrauensverhältnis zu ihrem:ihrer Teamteacher:in** berichteten, andere hingegen nicht. Es entstand der Eindruck, dass dieser Unterschied einen großen Einfluss auf die Wahrnehmung der Handlungen der Teamteacher:innen haben kann. Bisher existieren nur wenige Instrumente zur Messung des Vertrauens zwischen Lehrer:innen und Schüler:innen. Eine kleine Gruppe von Forscher:innen um Wayne Hoy hat sich jedoch intensiv mit dem Thema Vertrauen im schulischen Kontext befasst. Unsere Hauptquelle, die Skala Student Trust in Principal (STP) (z. B. in Forsyth et al., 2011, eigentlich 2004 entwickelt, Alpha: .95), entstammt dieser Arbeit. Sieben Items basieren auf der STP-Skala. Alle Items wurden auf die persönliche Beziehung zwischen Teamteacher:in und Schüler:in umformuliert, zwei Items wurden umgepolt. Da die Schüler:innen häufig von Erfahrungen mit Misstrauen oder enttäuschem Vertrauen berichteten, sollten die Schüler:innen die Möglichkeit erhalten, auch darüber Aussagen zu treffen (z.B. „Mein Teamteacher ist unfair zu mir.“). Drei weitere Items wurden neu hinzugefügt, welche Themen aufgriffen, die mehrfach in den Schüler:innen-Interviews genannt worden waren. Die ursprünglich 4-stufige Antwortskala (nie - immer) wurde umgekehrt und in eine 5-stufige Zustimmungsskala umgewandelt.

Im Kontrast zum Vertrauensverhältnis ist es auch möglich, dass Schüler:innen Situationen erleben, in denen sie sich von den Teamteacher:innen ungerecht behandelt fühlen. Zur **wahrgenommenen Fairness von Lehrpersonen** liegen verschiedene Skalen vor (z. B. Fend & Prester, 2014), in der PISA-I-Studie wurde beispielsweise ein Maß für „Beleidigung“ in die Bewertung einbezogen (Ramm et al., 2012). In der PISA-Studie 2015 wurde eine ähnliche, aber längere Skala mit einem stärkeren Fokus auf persönliche Erfahrungen verwendet (Kuger et al., 2016). Daraus wurden drei Items ausgewählt und angepasst, die nach persönlichen Erfahrungen mit der aktuellen Klasse und der Teamlehrkraft fragen (z.B. „Mein Teamteacher hat sich schon mal vor anderen Kindern über mich lustig gemacht.“). Ähnlich wie Fend und Prester (2014) wählten wir eine 5-Punkte-Skala für die Zustimmung, da Schätzungen der Häufigkeit für jüngere Schüler:innen besonders schwierig sein könnten. Parallel zu den drei Items über die **Teamteacher:innen** wurden auch Items zur **wahrgenommenen Fairness der Klassenleitungen** mit erhoben (z.B. „Mein Klassenlehrer hat mich schon mal angeschrien.“).

Ebenfalls Thema in den Schüler:innen-Interviews der ersten Förderperiode (siehe dazu Mitte et al., 2023, S. 32–33) war die **akademische Unterstützung durch den:die Teamteacher:in**. Aufbauend darauf wurde eine Skala mit neun selbst formulierten Items und 5-stufiger Zustimmungsskala entwickelt. Diese Skala erfasst die Meinungen der Schüler:innen zu den akademisch relevanten Aspekten, in denen der:die Teamteacher:in für sie einen Unterschied gemacht hat (z.B. „Durch den Teamteacher macht das Lernen mehr Spaß.“). Zweck der Skala ist es, herauszufinden, ob diese Meinungen universell oder auf bestimmte Klassen oder Schüler:innen beschränkt sind.

Anknüpfend an die SDT von Ryan und Deci (2000) wurden zu T1 zwei Skalen zur **Autonomieunterstützung durch die Lehrkraft bzw. den:die Teamteacher:in** aufgenommen. Die Items basieren auf der deutschen Übersetzung des Friendship Autonomy Support Questionnaire (Deci et al., 2006). Je sechs Items sind der Skala entnommen, wurden sprachlich leicht angepasst und auf den:die Klassenlehrer:in bzw. den:die Teamteacher:in bezogen (z.B. „Mein Klassenlehrer versteht mich.“) und mit einer 5-stufigen Zustimmungs-Skala versehen. Ab T2 wurde nur die Skala zu den Lehrkräften fortgeführt.

Bedürfniserfüllung, Wohlbefinden und Emotionsregulation

Wie bereits in Abschnitt 2.2.1 ausgeführt, definieren Ryan und Deci (2000) im Rahmen der SDT die Wichtigkeit dreier **psychologischer Grundbedürfnisse: Autonomie, Verbundenheit und Kompetenz**. Das Bestehen dieser drei Grundbedürfnisse trägt zum psychischen Wohlbefinden, der Selbstintegrität und der Motivation bei. Das motivationale Bedürfnis-Modell (motivational needs model) zeigt auf, dass soziale Interaktionen die individuelle Motivation einer Person beeinflussen, indem sie die drei Grundbedürfnisse entweder unterstützen und fördern oder diese einschränken (Furrer et al., 2014). Das Erleben von **Verbundenheit** in sozialen Interaktionen kann durch emotionale Wärme unterstützt oder durch Ablehnung beeinträchtigt werden. Das Erleben von **Kompetenz** wird durch eine klare Struktur unterstützt und entsprechend bedroht, wenn Chaos oder Willkür herrschen. Um jungen Menschen **Autonomie** zu ermöglichen, bedarf es Unterstützung, keinem Zwang oder Nötigung. Lehrpersonen und Mitschüler:innen tragen im schulischen Kontext dazu bei, die Bedingungen zur Erfüllung der Grundbedürfnisse zu gewährleisten (Alivernini & Lucidi, 2011; Furrer et al., 2014). Durch die Verwendung einer allgemeinen Bedürfnisskala für den Schulkontext können wir feststellen, wo sich die Kinder auf diesem Spektrum sehen. Zusätzlich untersuchten wir den Einfluss der Teamteachenden auf die Befriedigung der psychologischen Grundbedürfnisse ihrer Schüler:innen im Klassenraum. Wir verwendeten drei Subskalen, die ursprünglich von van den Broeck et al. (2010) entwickelt und von Kaiser et al. (2021) angepasst wurden. Die Autor:innen berichten ausreichende Reliabilität, d.h. Cronbach Alpha-Werte für Autonomie (5 Items, $\alpha = .78$),

Kompetenz (5 Items, $\alpha = .85$) und Verbundenheit (3 Items, $\alpha = .76$). Weil die Subskalen von Kaiser et al. (2021) allein die Bedürfnisbefriedigung im Biologieunterricht betrachten, änderten wir das Wording zu „während des Unterrichts“. Drei Items wurden umgekehrt kodiert. Die originale Antwortskala (0 = stimmt gar nicht, 4 = stimmt völlig) änderten wir zu 1= stimmt gar nicht, 5 = stimmt völlig. Weil bei einem Item während der T1-Erhebung mehrfach Nachfragen kamen („In der Schule fühle ich mich kompetent.“), wurde ab T2 dahinter in Klammern ergänzt „In der Schule kann ich Dinge gut.“. Zudem wurde ab T2 die Reihenfolge der Items so verändert, dass die drei Subskalen in sich geschlossen präsentiert wurden.

Es war uns außerdem ein Anliegen, das **Wohlbefinden der Schüler:innen im Schulalltag** genauer zu beurteilen. Akademische Emotionen sind ein wichtiger, wenn auch oft vernachlässigter Teil der Bildungsforschung. Zur genaueren Klassifikation akademischer Emotionen sowie einer Übersicht über den Forschungsstand, siehe z.B. Pekrun (2016). Reinhard Pekrun ist auch der Urheber der *Control-Value Theorie akademischer Emotionen*. Die Emotionen, die Schüler:innen empfinden, sind laut Pekrun (2016) abhängig von zwei Faktoren: (1) ihrer Kontrolle über die Situation („control appraisal“) und (2) der subjektiven Wichtigkeit („value appraisal“) der Situation. Adaptive Emotionen wie Stolz oder Lernfreude bilden die Grundlage von effektiver Selbstregulation im Unterricht. Maladaptive Emotionen wie Angst, Langeweile oder Hoffnungslosigkeit wirken sich hingegen negativ auf die akademischen Leistungen und die allgemeine psychische Gesundheit aus; dadurch bilden sie ein Risiko für einen Schulabbruch. Die Forschung weist diesbezüglich noch große Lücken auf und abgesehen von Prüfungsangst sind die Auswirkungen von Klassen-spezifischen Faktoren auf Emotionen der Schüler:innen noch kaum erforscht. Unser Fragebogen erhielt ein Item zum täglichen Wohlbefinden in der Schule, mit einer 11-stufigen Antwortskala (0 = schlechte Stimmung, 10 = sehr gute Stimmung). Die Skala war in Form eines Thermometers dargestellt, wobei der Tiefpunkt des Thermometers „schlechter Stimmung“ entsprach.

Verschiedene Strategien von **Emotionsregulation** stehen in engem Zusammenhang mit der persönlichen, sozialen und kognitiven Entwicklung (Gross, 2002). Wenn Emotionen beispielsweise zurückgehalten oder nicht nach außen gezeigt werden, wird von „Unterdrückung“ gesprochen. Neubewertung hingegen ist die Fähigkeit, eine Situation so zu interpretieren, dass sie einen anderen emotionalen Effekt erhält – indem z.B. eine einschüchternde Situation als positive Chance re-interpretiert wird. Vorteil einer Neubewertung ist, dass nicht nur das Aufkommen negativer Emotionen behindert wird, sondern auch bereits bestehende negative Emotionen „herunterreguliert“ werden (Gross & John, 2003). Die überwiegende oder ausschließliche Verwendung von Unterdrückung als Strategie der Emotionsregulation ist demnach ein Risikofaktor für sowohl die persönliche, soziale als auch die kognitive Entwicklung,

während Neuinterpretation eine adaptive Strategie darstellt. Um die Regulationsstrategien über die Zeit zwischen STT- und Kontrollgruppe vergleichen zu können, wurde zu allen Messzeitpunkten Emotionsunterdrückung und -neubewertung mit erhoben. Die verwendeten Items entstammen der deutschsprachigen Version von Abler und Kessler (2009) des Emotion-Regulation Questionnaire (ERQ) von Gross und John (2003). Die Items wurden kindgerechter gestaltet und die 7-Punkte-Antwortskala auf 5-Punkte verkürzt.

Schulerfolg

Schulabsentismus ist ein frühes Warnzeichen für verschiedene negative Entwicklungen (Kearney, 2008). Im Jahr 2019 betrug die Abbruchquote an Thüringer Schulen bei Schüler:innen 8.7% (Institut der Deutschen Wirtschaft Köln, 2019), weshalb Absentismus eine der zentralen Variablen der Erhebung darstellt. Besonders interessant für uns ist hierbei die absolute Häufigkeit der Fehltage. Da subjektive Angaben von Schüler:innen stark von der offiziellen Erhebung von Fehltagen abweichen können (Keppens et al., 2019), wurden parallel zur Erhebung im Rahmen der ÜNuF (siehe dazu **TEIL II – SCHÜLER:INNEN-BEFRAGUNG**) im Fragebogen die Häufigkeit geschwänzter Schulstunden und -tage abgefragt. Zuerst wurde abgefragt, ob der:die Schüler:in schon mal geschwänzt habe (Antwortskala: Ja, ich habe schon ein Mal geschwänzt/Nein, das mache ich nie.). Bei einem „Ja“ hatten die Schüler:innen die Möglichkeit, die Häufigkeit des Schwänzens innerhalb des Schuljahres anhand der vorgegebenen Häufigkeitskategorien (1 = ein Mal, 4 = mehrmals in der Woche) zu präzisieren oder die absolute Häufigkeit in einem offenen Textfeld selbst zu nennen. Diese offenen Angaben wurden im Nachhinein gesichtet und wenn möglich, in die vorgegebene Skala eingeordnet.

Als wichtige Indikatoren des Schulerfolgs wurden verschiedene Skalen zu **Engagement** bzw. zu dessen maladaptiven Gegenpart (**Disengagement**) im schulischen Kontext erhoben. In der angloamerikanischen Forschungsliteratur meint der Begriff *engagement* die aktive Beteiligung in einem sozialen Kontext (z. B. in der Schule), wobei „aktiv“ sehr weit ausgelegt werden kann. Insbesondere im Zusammenhang mit Schulabbruch stellt Engagement eine zentrale Prädiktorvariable dar (Gubbels et al., 2019). Üblicherweise wird Engagement als ein Metakonstrukt gesehen, das sich aus kognitiven, affektiven und verhaltensbezogenen Dimensionen zusammensetzt (für eine hilfreiche Übersicht siehe Fredricks & McColskey, 2012). Verhaltensengagement umfasst die Teilnahme an bestimmten Aktivitäten (z. B. Schularbeiten), insbesondere wenn diese über die täglichen Aufgaben hinausgehen (z. B. in außerschulischen Programmen), oder auch gutes Benehmen (z.B. pünktliches Erscheinen zum Unterricht). Bei kognitiv anspruchsvollen Verhaltensweisen wie Aufmerksamkeit, Ausdauer, Motivation und selbstgesteuertem Lernen finden sich uneinheitlich Angaben, welcher Facette diese zuzuordnen sind. Aspekte wie Ehrgeiz, „Wissensdurst“ und die Anwendung anspruchsvoller

Problemlösungsstrategien werden überwiegend „kognitivem“ Engagement zugeordnet. Obwohl Engagement in der Regel die Proaktivität der Schüler:innen umfasst, ist Disengagement nicht durch bloße Passivität gekennzeichnet. Stattdessen umfasst es aktiv maladaptive Zustände und Verhaltensweisen (Wang et al., 2019). Wir erfassen zwei separate Unterskalen zu Engagement und eine Disengagement-Skala: **kognitives Engagement** (nach Wang et al., 2019) und **agentisches Engagement** (nach Reeve & Tseng, 2011) sowie **verhaltensbezogenes Disengagement** (Wang et al., 2019). Alle Skalen wurden vom Englischen ins Deutsche übersetzt und mit einer 5-Punkt-Zustimmungsskala versehen. **Kognitives Engagement** wird mit fünf Items erfasst (z.B. „Ich strenge mich an, auch wenn ich mal mit meinen Aufgaben nicht weiterkomme.“), **agentisches Engagement**, d.h. *proaktives, selbstbestimmte Einbringen* in den Unterricht ebenfalls mit fünf Items (z.B. „Während des Unterrichts stelle ich Fragen.“) und verhaltensbezogenes **Disengagement** mit acht Items (z.B. „Ich mache im Unterricht Blödsinn.“).

Intrinsische Motivation bedeutet, etwas zu tun, weil die Handlung selbst Freude bereitet. **Extrinsische Motivation** hingegen bedeutet, etwas aus Zwang zu tun oder weil man sich eine Belohnung erhofft. Zwischen diesen beiden Polen gibt es eine Reihe weiterer „Motivationen“, die sich durch den Grad der Selbstbestimmung unterscheiden. Im Allgemeinen gilt: Je mehr man sich mit den Gründen seines Handelns identifiziert, desto intrinsischer ist die Motivation und desto selbstbestimmter handelt man. Es gibt viele wissenschaftliche Belege dafür, dass intrinsische Motivation nicht nur subjektiv ein „besseres Gefühl“ ist, sondern auch eine Reihe von Vorteilen hat, wie z. B. höhere schulische Leistungen (siehe z. B. Froiland et al., 2012 für eine Übersicht). Daher stellt sich oft die Frage, wie die intrinsische Motivation am besten gefördert werden kann. Nach Ryan und Deci (2000) ist die Befriedigung der drei Grundbedürfnisse nach Autonomie, Kompetenzerfahrung und Beziehung entscheidend. Viele Studien zeigen, dass ein autonomiefördernder Unterrichtsstil mit einer höheren intrinsischen Motivation der Studierenden verbunden ist (z.B. Núñez & León, 2015). Außerdem wurde eine dritte Dimension der Motivation hinzugefügt: die **identifizierte Motivation**. Diese beschreibt gemäß des SDT-Modells eine weitgehend selbstbestimmte extrinsischen Motivation, d.h. etwas macht mir zwar keinen Spaß, aber ich tue es, weil ich damit ein Ziel erreichen kann. Wir verwenden Items aus dem deutschen Fragebogen zur akademischen Selbstregulation für Jugendliche (SRQ-A [G]) von Gnambs und Hanfstingl (2014): jeweils vier Items zur intrinsischen (z.B. „Ich mache in der Schule mit, weil es mir Spaß macht.“) und extrinsischen Motivation (z.B. „...weil ich sonst Ärger mit den Lehrern bekomme.“ sowie drei Items zur identifizierten Motivation (z.B. „...weil ich so später einen besseren Job bekommen kann.“). Einige der Items wurden im Wortlaut vereinfacht und mit einer 5-Punkt-Zustimmungsskala kombiniert.

Mit realistischen **Bildungsaspirationen** geht eine subjektive Vorhersage über die tatsächlich möglichen Bildungswege und Abschlüsse einher, welche auf Kalkulationen der wahrgenommenen Chancen, Kosten und Nutzen basieren (Haller, 1968). Da eine Revision der einmal getroffenen Schulwahlentscheidung im deutschen Bildungssystem schwer ist, scheint es plausibel, dass Eltern die realistischen Abschlussziele, die sie für ihr Kind für erreichbar halten, am Übergang in die Sekundarstufe umsetzen. Demzufolge lassen sich realistische Aspirationen als vorgelagerte Bildungsentscheidungen interpretieren (Kleine, 2014). Die Operationalisierung der **idealistischen und realistischen Bildungsaspiration** ist der Operationalisierung der im Nationalen Bildungspanel aus der Bildungsberichterstattung (NEPS) (SC4) verwendeten Formulierung angelehnt und wurde für die in dieser Stichprobe vorliegende Altersgruppe sprachlich angepasst. Es wurde eine fünfstufige Antwortskala verwendet, mit 1 = kein Schulabschluss bis 5 = Universitätsabschluss. Der Universitätsabschluss stellt einen Ausbildungsabschluss dar, wurde hier allerdings als höchste Ausprägung einer möglichen Bildungszielsetzung verwendet.

Soziales Klassengefüge

Das **Klassenklima** hat eine starke Auswirkung auf das Verhalten der Schüler:innen, aber auch auf deren schulischen Erfolg (siehe z.B. MacAulay, 1990). Studien haben gezeigt, dass zwischen Schul- und Klassenklima und Schulabsentismus eine Verbindung besteht (siehe Sälzer, 2009). Zwei wichtige Parameter beim Untersuchen des Klassenklimas sind „Hilfsbereitschaft“ und „Rivalität“ (Jerusalem et al., 2009). Für die Erhebung nutzten wir die Skala „Kohäsion/Klassenzusammenhalt“, die bereits in der Studie „Hessischer Referenzrahmen Schulqualität“ (Institut für Qualitätsentwicklung Hessen, 2013) verwendet wurde, die ein Alpha von .75 für die 6-Item-Skala angaben. Wir nutzten 5 Items (z.B. „In unserer Klasse verstehen sich die meisten gut.“) und veränderten die vorher vierstufige Antwortskala (1 = stimmt gar nicht, 4 = stimmt genau) zu einer fünfstufigen Skala (1 = gar nicht, 5 = völlig).

Zusätzlich wurde das **(negative) Verhalten innerhalb der Klasse erfragt**. Dafür haben wir die Skala zur Messung der Disziplin von Wagner et al. (2011) verwendet. Die Skala richtet sich an Schüler:innen der weiterführenden Schule und einige der Items wurden bereits in der PISA-Studie 2000 verwendet. Wagner et al. (2011) geben Cronbachs Alpha mit .88 an. Die Items der Originalskala beziehen sich auf ein bestimmtes Unterrichtsfach, für unsere Erhebung haben wir diese Formulierung verallgemeinert, z.B. „Im Unterricht ist es laut und alles geht durcheinander.“. Außerdem wurde die vierstufige Antwortskala (1 = stimmt gar nicht, 4 = stimmt ganz genau) zu einer fünfstufigen Zustimmungsskala (1 = stimmt gar nicht, 5 = stimmt voll) verändert, die Formulierung leicht angepasst und ein Item entfernt.

Sozialer Hintergrund

Die Skala für **sozioökonomische Lernressourcen** basiert auf einem Maß, das wir in der ersten Phase des Forschungsprojekts verwendet haben. In einer Studie zu Disengagement und sozialer Identifikation erwies sich dieses Maß als ausreichender Proxy für den Sozioökonomischen Status (SES) (Glaesser et al., 2024). Die Grundidee war, ein **kindgerechtes Maß** zu schaffen, das nach der Verfügbarkeit von Raum und technischen Geräten zu Hause fragt, die mit dem sozioökonomischen Status der Familie in Verbindung gebracht werden können, wie z. B. ein eigenes Zimmer oder ein Computer, den die Kinder für ihre Hausaufgaben nutzen können. Ein Item stach in der Itemanalyse heraus und wurde daher entfernt („Zuhause gibt es einen Computer, den ich für Schulaufgaben benutzen kann.“).

Die **Bücherfrage** ist trotz fortgeschrittener Digitalisierungsprozesse bis heute ein weitgehend standardisiertes und validiertes Instrument, um Effekte der sozialen Herkunft zu operationalisieren (Heppt et al., 2022). Die Frage nach der im Haushalt vorhandenen Bücher bietet sich für diese Stichprobe insbesondere an, da diese auch valide von Kindern zu beantworten ist (ebd.). Der Einfluss der Bücheranzahl auf die Bildungsperformanz konnte wiederholt dargelegt werden (ebd.). Wir verwenden die Bücherfrage in der Version des **Thüringer Vergleichstests** (Nachtigall, 2008, 2018).

Demographischer Hintergrund

Bei der Erhebung des **demographischen Hintergrunds** wurde besonders Wert gelegt auf altersgerechte, geschlechtergerechte und diskriminierungssensible Items und Formulierungen, siehe dazu Abschnitt 1.6. **Alter** und das **Geburtsland** wurden zu Beginn des Fragebogens abgefragt, das **Geschlecht** sowie weitere Fragen zum **familiären Hintergrund** wurden zum Ende des Fragebogens erfragt, um durch ein Priming Antworttendenzen (bspw. aufgrund des Geschlechts) zu minimieren. Bei allen Items, die sich auf die Eltern bezogen, konnten die Schüler:innen zudem die Option auswählen „Kann oder will ich nicht beantworten“.

In Bezug auf die **Migrationsgeschichte** der Schüler:innen wurden mehrere Items verwendet, um einerseits die Diversität der individuellen Hintergründe abzubilden und andererseits kontextbezogen relevante Variablen in die Analyse einbeziehen zu können. So wurde neben dem **Geburtsland** der Schüler:innen auch einzeln die **Geburtsländer der Eltern** erfragt. Auf Basis dieser Antworten wurden mehrere Variablen erstellt, die beispielsweise einen Vergleich verschiedener Migrationsgenerationen ermöglichte. Zudem wurde die **Erstsprache** der Schüler:innen erfragt. Bei den Sprachen und Ländern waren zu T1 einige Auswahloptionen vorgegeben, in einem offenen Antwortfeld konnten die Schüler:innen nicht genannte Länder/Sprachen nennen, die rekodiert wurden und ab T2 in die Auswahloptionen übernommen wurden.

Bei der Angabe des **Geschlechts** konnten Schüler:innen den Satz „Ich bin...“ vervollständigen mit den Optionen „ein Junge“, „ein Mädchen“, „nicht-binär“, „trans“ oder unter „anderes, nämlich:“ selbst ihr Geschlecht angeben. Für das **Alter** wurden **Geburtsjahr** und -**monat** abgefragt, anhand derer über eine Verrechnung mit den genauen Erhebungszeitpunkten das Alter zu T1, T2 und T3 in die Analyse einbezogen werden konnte.

Einzelnen für beide Elternteile wurde erfragt, welchen **Abschluss** diese haben¹⁰. Der Bildungsabschluss der Eltern gilt als einer der stärksten Prädiktoren für die Bildungsleistung der Kinder. Neben den Einzelangaben zum jeweiligen Elternteil liegen zwei Variablen vor, welche entweder den gemittelten Bildungsabschluss beider Elternteile, bzw. den höchsten Bildungsabschluss beider Elternteile widerspiegeln.¹¹ Zudem wurde nach dem **aktuellen Beschäftigungsstatus** der Eltern gefragt (Vollzeit/Teilzeit/nicht berufstätig/kann oder will ich nicht beantworten). Der Beschäftigungsstatus der Eltern kann über verschiedene Mechanismen auf die Bildungsleistung und allgemeine Entwicklung der Schüler:innen einwirken. Grundlegend sind hierfür die Arbeiten von Bourdieu (1983) zu nennen, dessen Konzept der Kapitalausstattung zentral zur Erklärung von Bildungsdisparitäten ist. Es ist davon auszugehen, dass im Mittel Eltern, die sich in einem aktiven Erwerbsstatus befinden, über größere finanzielle Ressourcen verfügen, als Eltern, die über kein geregeltes Einkommen verfügen. Nach Bourdieu werden diese ökonomischen Ressourcen genutzt, um den Bildungserfolg der eigenen Kinder zu erhöhen (z.B. durch Anschaffung besserer Lernmaterialien). Weiterhin eröffnet das mit der Erwerbsarbeit zusammenhängende Prestige Zugang zu sozialen Ressourcen, die wiederum zur Erlangung von Bildungsvorteilen eingesetzt werden können (Bourdieu, 1983).

Die **Anzahl der Geschwister** kann über verschiedene Mechanismen auf die Bildungsleistung und allgemeine Entwicklung der Schüler:innen einwirken. Dabei ist die Effektrichtung nicht immer geklärt. Einerseits lässt sich hier wieder mit Bourdieu (1983) argumentieren: die begrenzte Ausstattung mit ökonomischem, sozialem und kulturellem Kapital muss auf mehrere Kinder aufgeteilt werden, was Kinder mit Geschwistern benachteiligt. Andererseits können ältere Geschwister, welche bereits eine höhere Schulform besuchen oder das Abitur erreicht haben, im Sinne einer Vorbildfunktion Aspiration und Motivation der jüngeren

¹⁰ Antwortoptionen: keinen Schulabschluss, Grundschule, Hauptschulabschluss (9. Klasse), Realschulabschluss (10. Klasse), Abitur (12. Klasse), Universitätsabschluss

¹¹ Jeder Abschlussoption wurde ein aufsteigender numerischer Wert zugeordnet. Liegen zu beiden Elternteilen Antworten vor, so wurde das Mittel beider numerischer Werte als mittlerer Bildungsabschluss herangezogen. Für den Fall, dass nur zu einem Elternteil der Bildungsabschluss angegeben wurde, wurde dieser als Mittel übernommen. Ähnlich wurde bzgl. des höchsten angegebenen Bildungsabschlusses der Eltern vorgegangen. Sollte zu beiden Elternteilen Angaben zum Abschluss vorliegen, wird der höchste Abschluss, bzw. im Falle eines gleich hohen Abschluss ebendieser übernommen. Liegen nur zu einem Elternteil Angaben vor, wurden lediglich diese übernommen. Liegen zu keinem Elternteil Abschlussinformationen vor, wurden beide Variablen als Missing gewertet.

Geschwister begünstigen (Grgic & Bayer, 2015). Dies gilt jedoch nicht, sofern die Unterstützungs- und Motivationsbereitschaft der Eltern lediglich auf dem Statuserhaltungsprinzip fußt und ältere Geschwister ebendiesen Statuserhalt bereits gewährleisten (Schulze & Preisendörfer, 2013). Die Schüler:innen konnten auswählen, ob sie keine, ein Geschwisterkind, zwei, drei oder mehr als drei Geschwister haben. Zusätzlich konnten sie angeben, ob ihre angegebenen Geschwister nicht, teilweise oder vollständig im selben Haushalt leben. Da einige Schüler:innen zum Wohnort ihrer Geschwister Angaben machten, obwohl sie zuvor angaben, keine Geschwister zu haben, wurde eine korrigierte Version dieser Variable angefertigt, welche Angaben zum Geschwisterwohnort auf „fehlend“ setzt, wenn zuvor keine Geschwister angegeben wurden.

Zudem wurden die Schüler:innen nach ihrer **Wohnsituation** gefragt. Dafür wurde zunächst gefragt, bei wem die Schüler:innen die meiste Zeit wohnen.¹² Ab T2 wurden weitere Antwortoptionen ergänzt (abwechselnd bei meiner Mutter und meinem Vater, bei meinen Pflegeeltern, bei meinen Adoptiveltern). Anschließend wurde gefragt, wo die Schüler:innen wohnen (in einer Wohnung, in einem Haus, in einer Gemeinschaftsunterkunft, in einem Jugendwohnheim oder einer Wohngruppe). Da Schüler:innen, die bspw. im Wechselmodell wohnen teilweise zwischen verschiedenen Wohnformen wechseln (z.B. in einem Haus bei der Mutter, in einer Wohnung beim Vater), wurde ab T2 die Anmerkung ergänzt „Wenn du mehrere Wohnorte hast, wähle den Ort aus, wo du am häufigsten bist.“.

2.3 Vergleichbarkeit zwischen Kontroll- und STT-Gruppe hinsichtlich zentraler soziodemographischer Variablen

Zunächst wurde die Vergleichbarkeit der Kontroll- und S -Gruppe hinsichtlich zentraler soziodemographischer Kennwerte zu T1 geprüft, eine wichtige Voraussetzung der vorliegenden Evaluation, sowohl für die Analyse der Schüler:innenfragebögen (Abschnitt 2.5) als auch der Übersichten der Noten und Fehltag (Abschnitt 3.3).

Sofern sich Kontroll- und STT-Gruppe hinsichtlich ihrer Schüler:innenzusammensetzung unterscheiden, ist eine Vergleichbarkeit dieser Klassen nicht ohne Kontrolle der sich unterscheidenden Charakteristika möglich. Neben Alter und Geschlecht zeigt bisherige Forschung wiederholt einen deutlichen Einfluss des sozioökonomischen Hintergrunds (SES) auf Bildungserfolg (Ditton & Maaz, 2022). Zur Spezifizierung des SES wurden die Schüler:innen wie oben beschrieben nach dem Bildungsabschluss, der Herkunft und Berufstätigkeit der Eltern, der Anzahl der Bücher und der Bildungs- und Lernressourcen zu Hause gefragt.

¹² Antwortoptionen: zusammen mit meinen Eltern, nur mit meiner Mutter zusammen, nur mit meinem Vater zusammen, bei anderen Verwandten, zusammen mit Freunden, allein, mit anderen Personen

In **Abbildung 1** und **Abbildung 2** sind Geschlecht und Alter gruppiert **nach ITT** abgebildet. Etwa 51.17% der Schüler:innen in der Kontrollgruppe und 52.27% der Schüler:innen der STT-Gruppe gaben ein männliches Geschlecht an. Unterteilt nach PP geben 50.35% der Schüler:innen in der Kontrollgruppe und 53.16% der Schüler:innen der STT-Gruppe ein männliches Geschlecht an. F-Test hat keine signifikanten Ergebnisse geliefert, die Verteilungen unterscheiden sich zwischen Gruppen also nicht signifikant. Das Durchschnittsalter in der STT-Gruppe zum ersten Erhebungszeitpunkt lag bei 12.11 (*SD*: 1.04), das der Kontrollgruppe bei 12.21 (*SD*: 1.10). Differenziert nach PP ergibt sich zum gleichen Zeitpunkt ein Durchschnittsalter in der STT-Gruppe von 12.03 (*SD* = 0.99) Jahren und in der Kontrollgruppe von 12.32 (*SD* = 1.14) Jahren. Der F-Test ($F = 9.819$, $p < .01$) deutet hier auf eine signifikant unterschiedliche Verteilung zwischen den Gruppen hin.

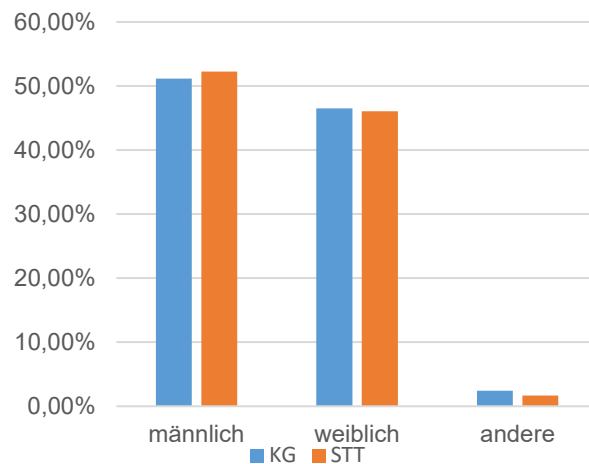


Abbildung 1 Verteilung der Geschlechterangaben nach ITT.

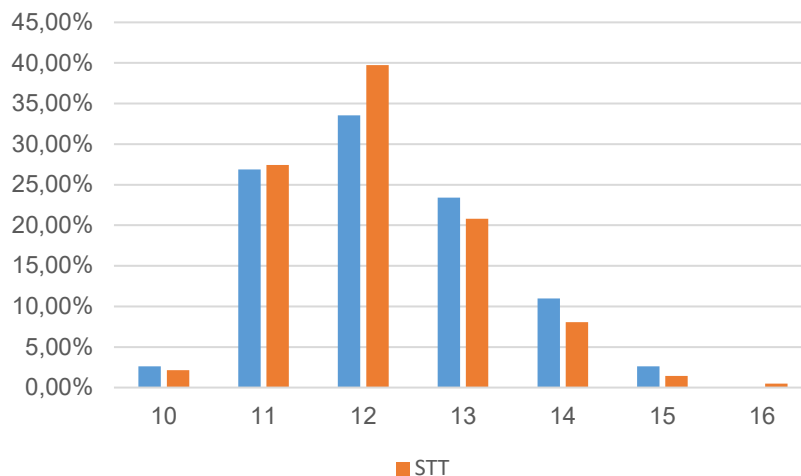


Abbildung 2 Verteilung der Altersangaben zum ersten Erhebungszeitpunkt nach ITT.

Unsere Analysen konnten zudem keine signifikanten Unterschiede in Bezug auf SES-relevante Variablen zwischen den Gruppen feststellen. So findet sich eine ähnliche Verteilung hinsichtlich der Angaben des höchsten Bildungsabschlusses der Eltern zwischen den beiden Gruppen nach ITT vor (**Abbildung 3**). Sofern nach PP differenziert wird, ist in der STT-Gruppe eine leichte Unterrepräsentation von Schüler:innen, deren Eltern maximal einen Realschulabschluss haben und eine leichte Überrepräsentation von Schüler:innen deren Eltern einen Hauptschulabschluss haben, festzustellen (**Abbildung 4**). Die Differenzen sind nicht statistisch signifikant. Ebenso wurden keine signifikanten Unterschiede hinsichtlich der Bücheranzahl (**Abbildung 4**) oder der Lernressourcen zu Hause (**Abbildung 5**) gefunden. Auch wenn nach PP zwischen den Gruppen differenziert wird, sind keine signifikanten Unterschiede zu finden, weshalb an diesen Stellen auf eine Abbildung der Differenzierung nach PP verzichtet wurde.

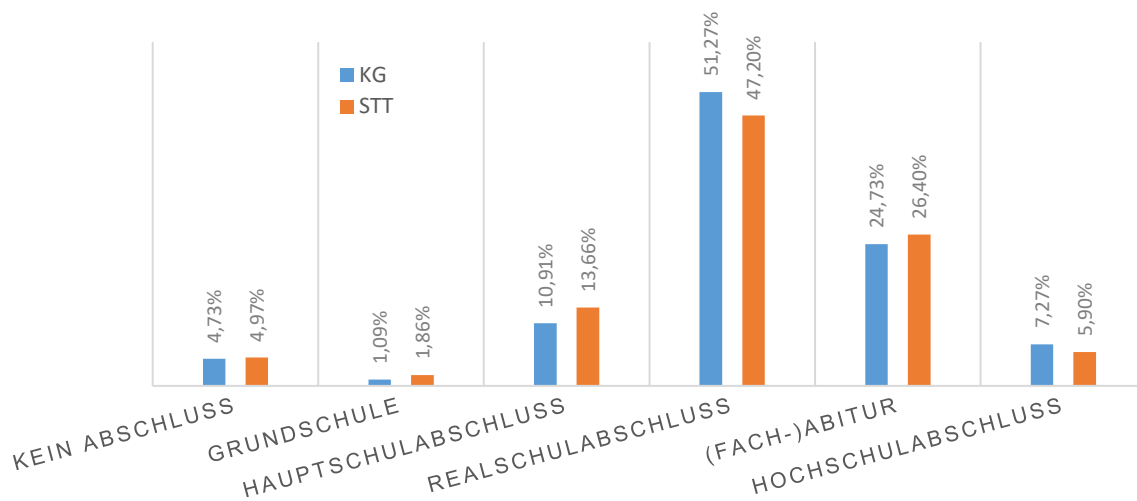


Abbildung 3 Verteilung des höchsten Bildungsabschlusses der Eltern nach ITT.

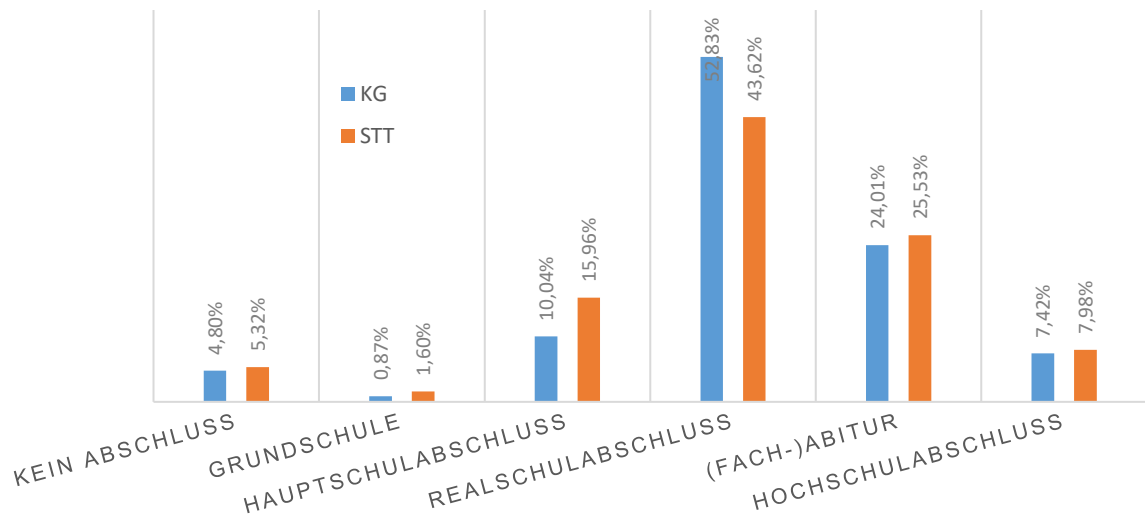


Abbildung 4 Verteilung des höchsten Bildungsabschlusses der Eltern nach PP.

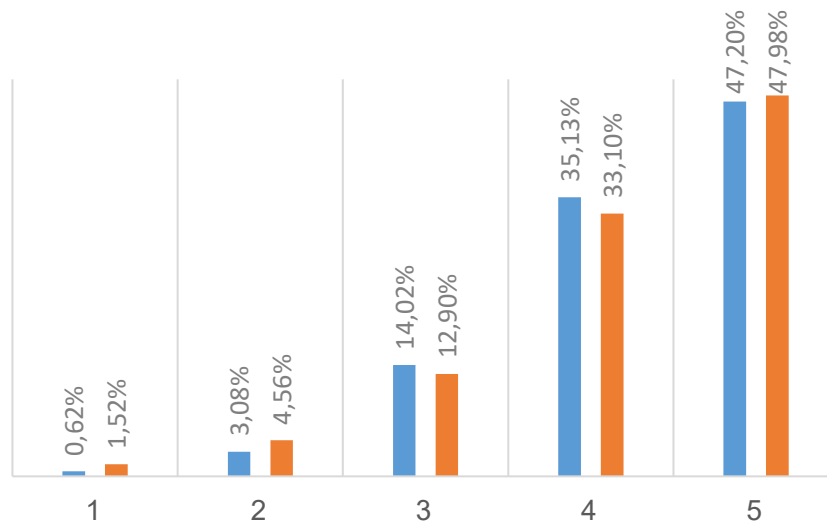


Abbildung 5 Verteilung der Angaben bzgl. der zu Hause zur Verfügung stehenden Lernressourcen nach ITT.

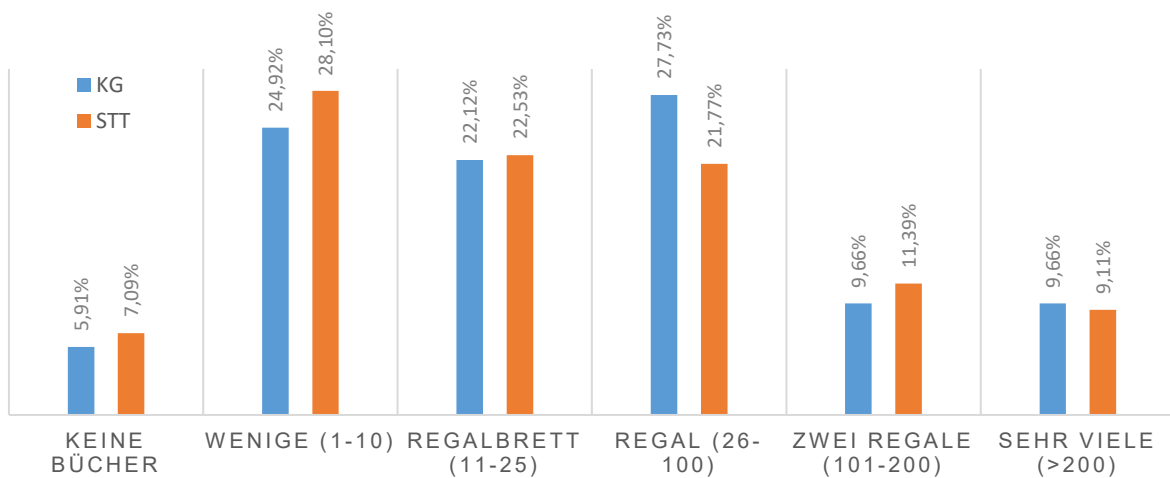


Abbildung 6 Verteilung der Bücheranzahl zu Hause nach ITT.

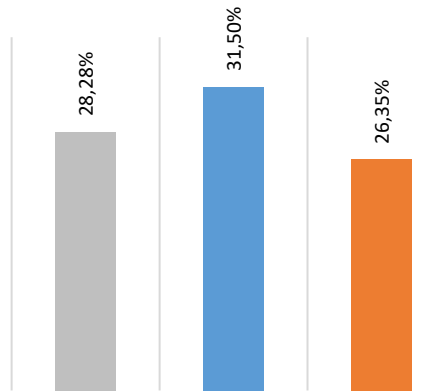


Abbildung 7 Anteil SuS mit Migrationshintergrund insgesamt (grau), in Kontrollgruppe (blau) und STT-Gruppe (orange).

Hinsichtlich des Migrationshintergrunds und der gelernten Erstsprache wurden, sofern nach ITT differenziert wurde, Unterschiede zwischen den Gruppen festgestellt. Unter die Kategorie Migrationshintergrund wurden alle Schüler:innen zusammengefasst, die selbst außerhalb Deutschlands geboren waren oder angaben, mindestens einen im Ausland geborenen Elternteil zu haben. So beträgt der Anteil der Schüler:innen mit Migrationshintergrund in der STT-Gruppe 26.35% und in der Kontrollgruppe 31.5%. Die Differenz ist gemessen am χ^2 -Test allerdings nicht signifikant. Wenn nach PP unterschieden wird, sind die Diskrepanzen geringer (STT: 32.08 % vs. KG: 31.14%). Bei der Betrachtung der erlernten Erstsprache lassen sich signifikante Unterschiede bei ITT-Differenzierung feststellen: Während 22.19% der Schüler:innen der STT-Gruppe angaben, eine andere Sprache als deutsch zuerst gelernt zu haben, gaben dies 28.66% der Schüler:innen in der Kontrollgruppe an ($\chi^2(1, N = 722) = 3.96, p = .05$). Auch hier lassen sich allerdings keine signifikanten Differenzen feststellen, wenn nach PP unterschieden wird (STT: 28.19% vs. KG: 27.92%).

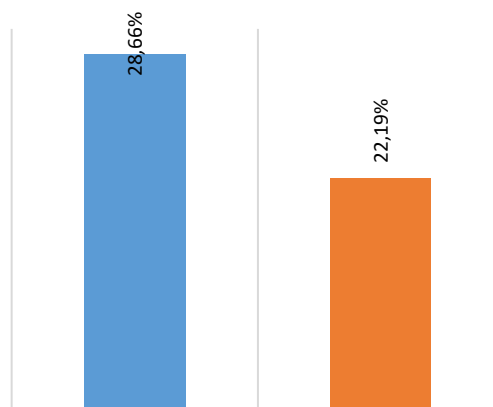


Abbildung 8 Anteil SuS mit nicht-deutscher Erstsprache in Kontrollgruppe (blau) und STT-Gruppe (orange).

Evaluationsrelevante Unterschiede zwischen den Gruppen sind über Disparitäten hinsichtlich des SES hinaus denkbar. Hierfür wurden die Teamteachenden gefragt, ob ihre Klasse spezifisch für das STT zusammengestellt wurde. Auch wurden die Schulleitungen gefragt, ob Klassen aufgrund bereits vorliegender Unterschiede STT spezifisch zugewiesen wurde. Diese Angaben wurden als neue Variable in den Datensatz aufgenommen und in die Analysen einbezogen (siehe 5.3.2), um ggf. grundlegende Unterschiede in den Klassen zu berücksichtigen. Es kann hier schon vorweggenommen werden, dass kein signifikanter Einfluss dieser Variablen erkennbar war.

2.4 Methodik der Datenanalyse

Kern des Ergebnisteils bildet die Analyse der Längsschnittdaten in den folgenden Abschnitten. Diese wurden überwiegend anhand von Zweifaktoriellen Varianzanalysen (ANOVA) mit Messwiederholungen ausgewertet, welche einen Vergleich zwischen Kontroll- und STT-Gruppe über die drei Messzeitpunkte hinweg ermöglichen. Diese Analysen schließen jeweils eine abhängige Variable mit ein. Wenn andere Analysemodelle angewendet wurden, werden diese kurz erläutert.

ANOVA ermöglichen, **Veränderungen über die Zeit** sowie **Unterschiede zwischen Gruppen** (in unserem Fall: Interventions- vs. Kontrollgruppe) in einem gemeinsamen Modell zu betrachten. So wird das Risiko verringert, dass zufällige Unterschiede fälschlicherweise als bedeutsam interpretiert werden. Ein **signifikanter Effekt** bedeutet in diesem Zusammenhang, dass ein beobachteter Unterschied zwischen Gruppen oder eine Veränderung über die Zeit mit hoher Wahrscheinlichkeit auf die Intervention zurückführen kann, weil es statistisch sehr unwahrscheinlich ist, dass er durch Zufall entstanden ist. Die Signifikanz eines Effekts ist am **p-Wert** abzulesen, wenn dieser niedriger als **.05** ist, gilt der Effekt als signifikant. Von einem **marginal signifikanten Effekt** wird gesprochen, wenn der p -Wert $< .10$ ist.

Die **Effektgröße** beschreibt die **Stärke eines Unterschieds** unabhängig von der Stichprobengröße. Effekte können positiv (der Wert der abhängigen Variable nimmt über die Zeit zu) oder negativ (der Wert nimmt über die Zeit ab) sein. Neben Haupteffekten der Zeit (z.B. es gibt eine Abnahme von Motivation über die Zeit) und der Interventionsbedingung (z.B. Klassen mit STT sind niedriger motiviert als Klassen ohne STT) können **Interaktionseffekte** auftreten. Das bedeutet, dass sich der Verlauf einer Veränderung **zwischen den Gruppen unterscheidet** (z.B. die Motivation in den Klassen mit STT nimmt über die Zeit stärker ab als in den Klassen ohne STT). Ein signifikanter Interaktionseffekt zeigt also, dass sich die Klassen mit STT im Vergleich zur Kontrollgruppe **unterschiedlich entwickeln**.

Im schulischen Kontext sind die zu erwartenden Effektgrößen von Interventionen in der Regel eher klein bis mittelgroß (Murawski & Swanson, 2001), da Lernprozesse und Verhaltensänderungen von vielen verschiedenen Faktoren beeinflusst werden. In den hier verwendeten ANOVAs wird der Wert des **partiellen Eta-Quadrat (η_p^2)** als Maß für die Effektstärke angegeben, wobei η_p^2 zeigt, wie stark sich eine Maßnahme im Verhältnis zu allen anderen Einflussfaktoren bemerkbar macht: Ein Wert von etwa **0.01** gilt als **kleiner Effekt**, etwa **0.06** als **mittlerer Effekt** und ab ungefähr **0.14** als **großer Effekt**.

Ob ein Effekt jedoch **statistisch signifikant** wird, hängt auch von der Stichprobengröße ab. Bei großen Stichproben ist die **Power** höher, d.h., es können bereits sehr kleine Effekte als signifikant erscheinen, während bei kleinen Stichproben selbst größere Effekte statistisch unauffällig bleiben können. Eine ausreichende Stichprobengröße erhöht somit die Robustheit und Verlässlichkeit des Modells und trägt dazu bei, dass tatsächliche Effekte erkannt und Zufallsergebnisse minimiert werden. Wenn beispielsweise die Stichprobengröße reduziert wird und die Effekte in derselben Richtung erhalten bleiben, spricht dies für die **Robustheit** des Effekts.

Um die im Abschnitt 2.1 beschriebenen Veränderungen der Stichprobe hinsichtlich STT- und Kontrollgruppe auch in der Analyse zu berücksichtigen, wurden getrennte Analysen für **intention to treat (ITT)** und **per protocol (PP)** gerechnet. Dabei handelt es sich nicht um separate Stichproben, sondern um unterschiedliche Kriterien für die Aufnahme in die Analyse. Da die Auswahlkriterien für die ITT-Bedingung weniger streng sind, können in diese Analysen mehr Klassen aufgenommen werden, was wiederum die Power erhöht. In der aufgrund der strengeren Auswahlkriterien des PP-Ansatzes reduzierten Stichprobengröße ist die Power in diesen Analysen niedriger, sodass ggf. kleinere Effekte nicht signifikant werden.

Bei der Interpretation der Ergebnisse ist zusätzlich zu beachten, dass nur jene Variablen in das Modell einbezogen werden können, die tatsächlich erhoben wurden. Weitere Faktoren, die möglicherweise einen Einfluss auf die Ergebnisse haben (z. B. situative Bedingungen oder der Einfluss des Elternhauses), bleiben unberücksichtigt, wenn sie nicht gemessen wurden. Daher sollte die **Interpretation der Ergebnisse stets im Kontext der erhobenen Daten** erfolgen. Einen wichtigen Beitrag liefert hier die Vergleichbarkeit der Kontroll- und Interventionsklassen hinsichtlich der soziodemographischen Zusammensetzung, siehe dazu Abschnitt 2.3. Wo sich signifikante Unterschiede zwischen Interventions- und Kontrollgruppe fanden, wurden die entsprechenden soziodemographischen Variablen in die Modelle aufgenommen, von denen angenommen wurde, dass sie – zusätzlich zur Intervention – die Veränderung der abhängigen Variable über die Zeit beeinflussen könnten (siehe dazu Abschnitt 2.3).

In den folgenden Abschnitten wird berichtet, ob und inwiefern die Implementation des STT Auswirkungen über die Zeit auf die Wahrnehmung des STT (2.5.1), Bedürfnisbefriedigung und psychologisches Wohlbefinden (2.5.2), Schulerfolg (2.5.3) sowie das soziale Klassengefüge (2.5.4) hat. Im Abschnitt 2.6 werden die zentralen Ergebnisse zusammengeführt und diskutiert.

2.5 Ergebnisse der Schüler:innenfragebögen

2.5.1 Skalen zum STT und zu Lehrpersonen

Vertrauensverhältnis zum:zur Teamteacher:in

Das Vertrauensverhältnis zum:zur Teamteacher:in war eine der zentralen Variablen der ersten Förderperiode. In dieser Analyse wurde nur die STT-Gruppe nach ITT- oder PP-Aufteilung inkludiert ($n_{ITT} = 131$, $n_{PP} = 111$). Einfaktorielle ANOVAs mit Messwiederholung (ohne Gruppenunterscheidung) zeigten signifikante Haupteffekte für den zeitlichen Verlauf, $F_{ITT}(2, 260) = 5.833$, $p < .01$, $\eta_p^2 = .04$; $F_{PP}(2, 220) = 9.122$, $p < .001$, $\eta_p^2 = .08$. Das Vertrauen war zu Beginn hoch ($M_{T1,ITT} = 4.32$, $M_{T1,PP} = 4.36$) und nahm über die Zeit leicht ab ($M_{T2,ITT} = 4.15$, $M_{T2,PP} = 4.12$; $M_{T3,ITT} = 4.09$, $M_{T3,PP} = 4.05$). Dieser Verlauf gestaltete sich gleichmäßig, wie in **Abbildung 9** einzusehen ist, $F_{ITT}(1,130) = 9.907$, $p < .01$, $\eta_p^2 = .07$; $F_{PP}(1,110) = 16.165$, $p < .001$, $\eta_p^2 = .13$.

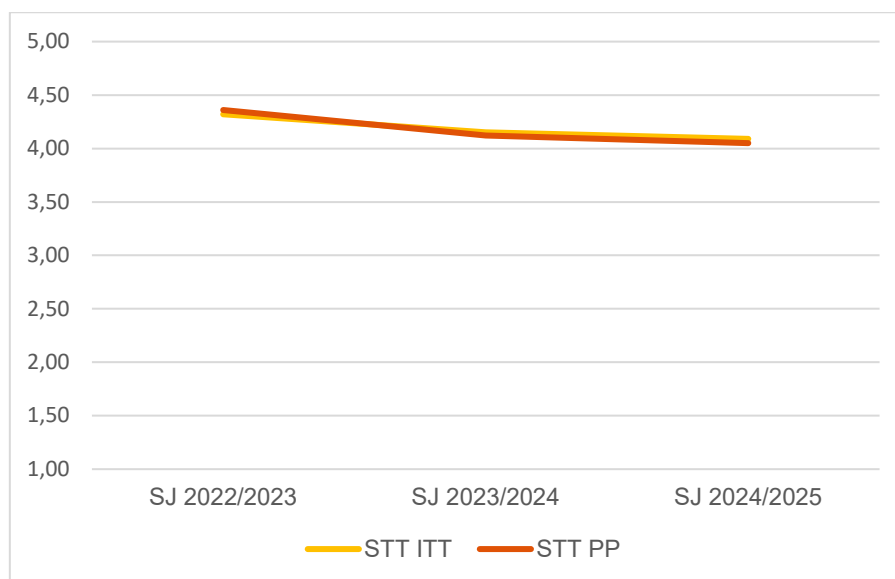


Abbildung 9 Zeitlicher Verlauf von Vertrauen der SuS in Teamteachende.

Wahrgenommene Fairness von Teamteachenden

Zur Fairness der Teamteachenden wurde für die STT-Gruppe ($n_{STT} = 131$, $n_{PP} = 110$) eine einfaktorielle ANOVA berechnet. Es ergab sich keine signifikante Veränderung über die Zeit (siehe **Abbildung 10**), die wahrgenommene Fairness wird als sehr hoch eingeschätzt. Auch in der Aufteilung nach PP findet sich keine signifikante Veränderung über die Zeit.

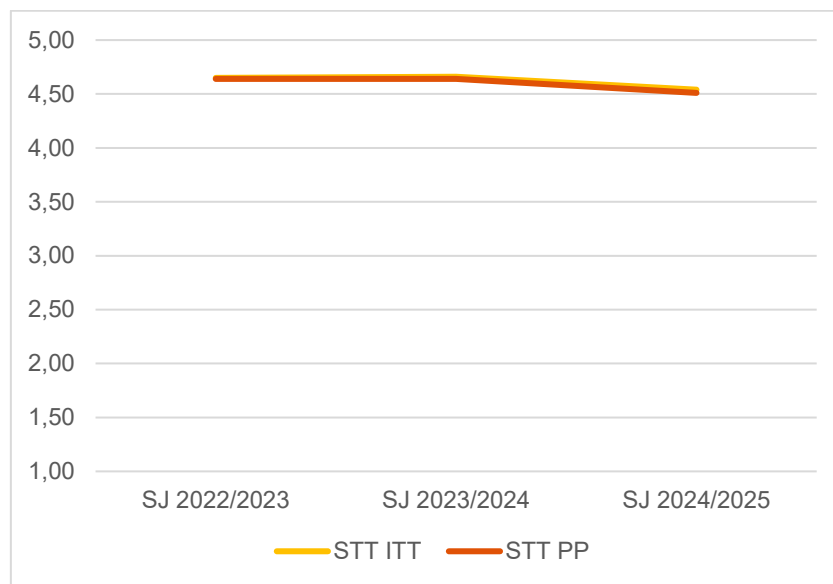


Abbildung 10 Zeitlicher Verlauf von Fairness der Teamteachenden.

Wahrgenommene Fairness von Klassenlehrkraft

Nach der ITT-Gruppeneinteilung zeigte sich, dass sich die STT- und Kontrollgruppe über die Zeit hinsichtlich der wahrgenommenen Fairness der Klassenlehrkraft unterschieden, $F_{ITT}(2, 446) = 5.574, p < .01, \eta_p^2 = .01$. Während sich die Entwicklung der wahrgenommenen Fairness bei Schüler:innen der Kontrollgruppe zu T3 erhöhte ($M_{T1} = 4.24, M_{T2} = 4.05, M_{T3} = 4.16$), sank sie bei den Schüler:innen der STT-Gruppe ab ($M_{T1} = 4.11, M_{T2} = 4.10, M_{T3} = 3.91$; siehe **Abbildung 11**). Außerdem ergab sich nach dieser Gruppeneinteilung ein signifikanter Haupteffekt der Zeit, $F(2, 894) = 5.222, p < .01, \eta_p^2 = .01$, welcher linear verlief, $F_{ITT}(1, 447) = 8.547, p < .01, \eta_p^2 = .02$. Bei einer Aufteilung nach PP und somit einer verminderten Stichprobengröße konnten diese Effekte nicht festgestellt werden, d.h. die Verläufe für STT vs. KG sind gleich.

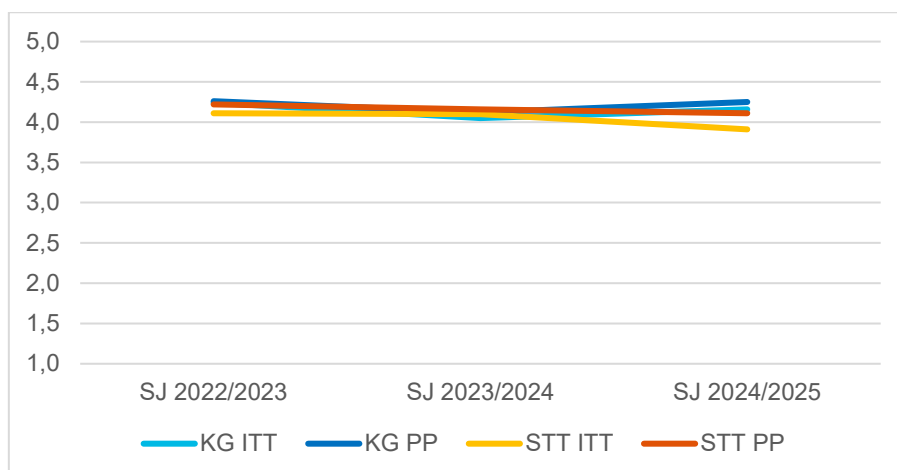


Abbildung 11 Zeitlicher Verlauf der wahrgenommenen Fairness der Klassenlehrkraft.

Akademische Unterstützung durch Teamteacher:in

Für die wahrgenommene akademische Unterstützung durch den:die Teamteacher:in wurde ebenfalls für die STT-Gruppe eine einfaktorielle ANOVA berechnet. In den deskriptiven Werten ($n_{\text{STT}} = 132$) ist eine Abnahme der wahrgenommenen Unterstützung zu T2 erkennbar, zu T3 stabilisiert sich der Wert ($M_{T1} = 3.39$, $M_{T2} = 3.25$, $M_{T3} = 3.26$). Es zeigt sich jedoch kein signifikanter Effekt über die Zeit hinweg. In der leicht reduzierten Stichprobengröße der Aufteilung nach PP zeigt sich deskriptiv ein ähnliches Muster ($M_{T1} = 3.48$, $M_{T2} = 3.28$, $M_{T3} = 3.29$). Hier findet sich jedoch ein signifikanter Effekt für die Zeit ($F(2, 218) = 4.01$, $p < .05$, $\eta_p^2 = .04$), ersichtlich in einem linearen Abwärtstrend ($F(1, 109) = 5.29$, $p < .05$, $\eta_p^2 = .05$).

Wahrgenommene Autonomieunterstützung durch die Lehrkraft

Die zeitliche Entwicklung der Mittelwerte der wahrgenommenen Autonomieunterstützung durch die Lehrkraft kann in **Tabelle 5** Mittelwerte und Standardabweichungen der wahrgenommenen Autonomieunterstützung durch die Lehrkraft eingesehen werden. Unabhängig davon, ob nach ITT oder PP differenziert wird, sinkt die wahrgenommene Autonomieunterstützung in der STT-Gruppe, während sie in der Kontrollgruppe auf dem gleichen Niveau verbleibt, bzw. leicht ansteigt. Insofern ist ein signifikanter Interaktionseffekt mit der Zeit für ITT- ($F(2, 906) = 8.99$, $p < .001$, $\eta_p^2 = .019$) und PP-Differenzierung ($F(2, 616) = 3.8$, $p < .05$, $\eta_p^2 = .012$) festzustellen. Auch der Verlauf in der STT-Gruppe unterscheidet sich signifikant von der Kontrollgruppe (STT: $F(1, 453) = 14.2$, $p < .001$, $\eta_p^2 = .03$, PP: $F(1, 308) = 6.26$, $p < .05$, $\eta_p^2 = .02$). Es ist ein signifikanter Haupteffekt über die Gruppen sowohl bei ITT ($F(1, 453) = 6.49$, $p < .05$, $\eta_p^2 = .014$) als auch PP ($F(1, 308) = 5.8$, $p < .05$, $\eta_p^2 = .018$) zu berichten.

Tabelle 5 Mittelwerte und Standardabweichungen der wahrgenommenen Autonomieunterstützung durch die Lehrkraft

				T1		T2		T3		
				<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Wahrgenommene Autonomieunterstützung durch die Lehrkraft	ITT	STT	244	3.47	0.94	3.31	0.98	3.15	1.07	
		KG	211	3.52	0.98	3.44	0.97	3.57	1.05	
	PP	STT	175	3.47	0.96	3.31	0.98	3.26	0.97	
		KG	135	3.55	0.96	3.48	0.95	3.65	0.99	

2.5.2 Bedürfniserfüllung, Wohlbefinden und Emotionsregulation

Psychologische Grundbedürfnisse: Verbundenheit, Kompetenz und Autonomie

Die zeitliche Entwicklung der psychologischen Grundbedürfnisse Verbundenheit, Kompetenzerleben und Autonomie ist in **Tabelle 6** dargestellt. In der Gruppenaufteilung nach ITT finden sich sowohl bei der Erfüllung des Bedürfnisses nach Verbundenheit als auch nach

Autonomie signifikante Interaktionseffekte zwischen der Zeit und der Interventionsgruppe (Verbundenheit: $F(2, 920) = 11.64, p < .001, \eta_p^2 = .03$; Autonomie: $F(2, 926) = 3.18, p < .05, \eta_p^2 = .01$). Das Erleben von Verbundenheit ist anfänglich größer in der STT-Gruppe, nimmt jedoch leicht ab, während es in der Kontrollgruppe über die Zeit zunimmt ($F(1, 460) = 20.88, p < .001, \eta_p^2 = .04$). Das wahrgenommene Autonomieerleben ist anfänglich gleich hoch und sinkt dann in beiden Gruppen über die Zeit ab, jedoch stärker in der STT-Gruppe ($F(1, 463) = 4.59, p < .05, \eta_p^2 = .01$). Im zeitlichen Verlauf des Autonomieerlebens findet sich zudem ein Haupteffekt für die Zeit ($F(2, 926) = 22.89, p < .001, \eta_p^2 = .05$), der Wert nimmt über die Zeit ab ($F(1, 463) = 38.62, p < .001, \eta_p^2 = .08$). Für das Kompetenzerleben findet sich ein signifikanter Haupteffekt für die Gruppe, die Kontrollgruppe berichtet durchgehend höheres Kompetenzerleben als die STT-Gruppe ($F(1, 463) = 7.79, p < .05, \eta_p^2 = .02$).

In der Aufteilung nach PP finden sich trotz erheblich reduzierter Stichprobengröße ähnliche Muster und dieselben Effekte. Beim Erleben von Verbundenheit findet sich ein signifikanter Interaktionseffekt ($F(2, 620) = 13.12, p < .001, \eta_p^2 = .04$): anfänglich ist das Erleben von Verbundenheit in der STT-Gruppe noch höher als in der ITT-Gruppierung und es nimmt in der Kontrollgruppe zwar erst zu T3 zu, dennoch findet sich auch hier ein linearer Interaktionseffekt ($F(1, 310) = 24.48, p < .001, \eta_p^2 = .07$). Beim Autonomieerleben liegen ein Haupteffekt für die Zeit ($F(2, 624) = 20.43, p < .001, \eta_p^2 = .06$) sowie eine Interaktion vor ($F(2, 624) = 4.96, p < .01, \eta_p^2 = .02$): in beiden Gruppen nimmt der Wert über die Zeit ab ($F(1, 312) = 32.86, p < .001, \eta_p^2 = .10$). Anfänglich liegt der Wert für T1 in der STT-Gruppe höher als in der ITT-Gruppierung, sinkt jedoch über die Zeit stärker ab als in der Kontrollgruppe ($F(1, 312) = 6.52, p < .05, \eta_p^2 = .02$). Auch der Haupteffekt für den Gruppenunterschied beim Kompetenzerleben zeigt sich in der reduzierten Stichprobe ($F(1, 312) = 4.75, p < .05, \eta_p^2 = .02$).

Tabelle 6 Mittelwerte und Standardabweichungen zu den psychologischen Grundbedürfnissen

				T1		T2		T3	
				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>N</i>									
Verbundenheit	ITT	STT	244	3.54	1.01	3.25	1.05	3.25	0.96
		KG	218	3.31	1.03	3.34	1.00	3.47	1.00
	PP	STT	136	3.63	0.95	3.37	1.07	3.23	0.96
		KG	176	3.31	0.99	3.31	0.99	3.49	1.01
Kompetenzerleben	ITT	STT	246	3.49	0.68	3.37	0.74	3.39	0.72
		KG	219	3.59	0.63	3.53	0.67	3.58	0.69
	PP	STT	138	3.48	0.64	3.33	0.74	3.37	0.73
		KG	176	3.53	0.61	3.49	0.69	3.58	0.72
Autonomie	ITT	STT	246	3.21	0.69	2.99	0.65	2.92	0.63
		KG	219	3.21	0.70	3.12	0.67	3.07	0.67
	PP	STT	138	3.28	0.69	2.97	0.62	2.92	0.66

Wohlbefinden der Schüler:innen im Schulalltag

Beim Wohlbefinden im Schulalltag zeigt sich in der Gruppierung nach ITT ein signifikanter Haupteffekt für die Zeit ($F(2, 900) = 9.53, p < .001, \eta_p^2 = .02$) sowie ein Interaktionseffekt ($F(2, 900) = 4.67, p < .001, \eta_p^2 = .01$). Allgemein findet sich eine lineare Abnahme über die Zeit ($F(1, 450) = 17.28, p < .001, \eta_p^2 = .04$). Während das Wohlbefinden in der STT-Gruppe anfangs höher ist, sinkt es über die Zeit ab, siehe **Abbildung 12** ($n = 238$). In der Kontrollgruppe gibt es von T1 zu T2 ebenfalls eine Abnahme, jedoch zu T3 wieder eine Zunahme ($n = 214$), resultierend in einem linearen Interaktionseffekt ($F(1, 450) = 9.42, p < .01, \eta_p^2 = .02$). In der reduzierten Stichprobe der PP-Gruppierung finden sich diese Effekte ebenfalls, die Effektstärken nehmen sogar zu (Zeit: $F(2, 616) = 13.65, p < .001, \eta_p^2 = .04$), Interaktion: $F(2, 616) = 7.26, p < .001, \eta_p^2 = .02$). Das Wohlbefinden nimmt über die Zeit ab ($F(1, 302) = 26.91, p < .001, \eta_p^2 = .08$). Im Vergleich der Gruppen zeigt sich erneut ein Interaktionseffekt ($F(1, 302) = 15.16, p < .001, \eta_p^2 = .05$): die Schüler:innen der STT-Gruppe berichten anfänglich höheres Wohlbefinden, dieses sinkt jedoch über die Zeit noch deutlicher ($n = 133$). In der Kontrollgruppe ist die Entwicklung ebenfalls vergleichbar gegenüber der größeren ITT-Gruppierung ($n = 171$).

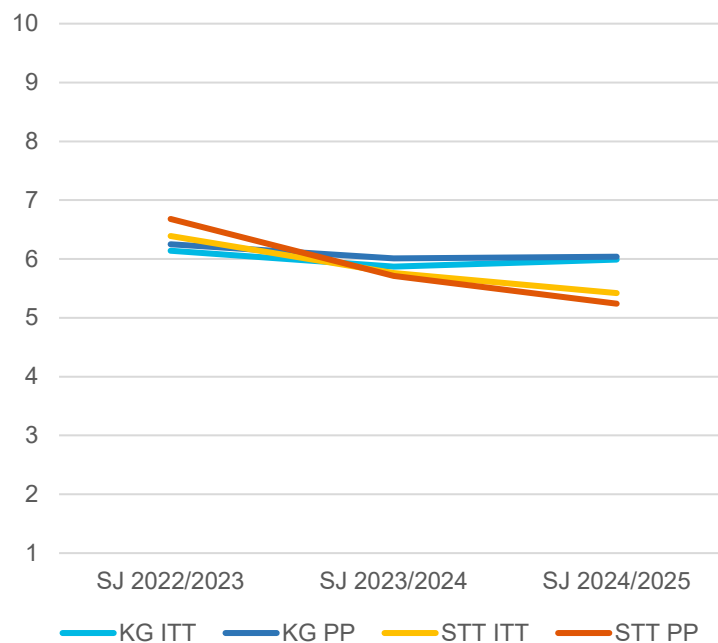


Abbildung 12 Zeitliche Entwicklung des Wohlbefindens im Schulalltag, 1 = negatives und 10 = positives Befinden.

Emotionsregulation: Neubewertung und Unterdrückung

Ausgehend von dem STT-Manual (Kindersprachbrücke Jena e. V., 2019) wurde angenommen, dass sich die Verantwortlichkeit der Teamteachenden für sozial-emotionale Kompetenzen auch positiv auf die Emotionsregulation der Schüler:innen auswirken könnte. Sowohl in der ITT- als auch in der PP-Aufteilung ließen sich jedoch hier keine statistisch signifikanten Effekte zeigen. In beiden Aufteilungen unterschieden sich die Werte kaum zwischen der Kontroll- und STT-Gruppe, auch über die Zeit schwankten die Werte nur leicht, siehe dazu **Tabelle 7**. In der ITT-Bedingung zeigten sich lediglich marginal signifikante Interaktionseffekte für Neubewertung ($F(2, 804) = 3.01, p < .10, \eta_p^2 = .01$) und Unterdrückung ($F(2, 802) = 2.88, p < .10, \eta_p^2 = .01$). In der PP-Bedingung traten keine signifikanten Effekte auf.

Tabelle 7 Mittelwerte und Standardabweichung für Emotionsregulation

Strategie			<i>n</i>	T1 <i>M (SD)</i>	T2 <i>M (SD)</i>	T3 <i>M (SD)</i>
Neubewertung	ITT	STT	219	3.08 (0.84)	2.96 (0.78)	3.04 (0.79)
		KG	185	3.09 (0.81)	3.14 (0.83)	3.03 (0.75)
	PP	STT	126	3.08 (0.80)	2.96 (0.81)	3.04 (0.85)
		KG	155	3.09 (0.84)	3.12 (0.82)	3.05 (0.77)
Unterdrückung	ITT	STT	218	3.18 (0.86)	3.08 (0.84)	3.28 (0.82)
		KG	185	3.17 (0.86)	3.20 (0.87)	3.17 (0.83)
	PP	STT	126	3.19 (0.84)	3.15 (0.86)	3.26 (0.85)
		KG	155	3.17 (0.88)	3.19 (0.87)	3.15 (0.84)

2.5.3 Schulerfolg

Schulabsentismus

Schulabsentismus, also das unentschuldigte Fernbleiben vom Unterricht, wurde grundsätzlich über zwei verschiedene Formen der Abfrage operationalisiert. Zum einen wurde mithilfe der Übersichten über Noten und Fehltage die konkrete Anzahl der unentschuldigten und entschuldigten Fehltage (wie im Halb- bzw. Endjahreszeugnis dokumentiert) abgefragt. Diesbezügliche Analysen sind im Abschnitt 3.3.1 zu finden. Zum anderen wurden die Schüler:innen direkt zu ihrem Absentismusverhalten befragt. Zunächst wurden die Schüler:innen gefragt, ob sie jemals geschwänzt haben. Wenn hierauf eine positive Antwort gegeben wurde, wurde weiterführend nach der Häufigkeit des Schwänzens in vier Kategorien (ein Mal bis mehrmals die Woche, siehe auch **Abbildung 13**) gefragt sowie die Option gegeben, die Häufigkeit in eigenen Worten zu beschreiben. Die in den fortfolgenden Analysen verwendeten Daten basieren auf der Zusammenführung der offenen Angaben und der vier vorgegebenen Antwortkategorien. Sollten offene Antworten nicht kategorisiert werden können, wurden sie in der Kategorie NA gezählt.

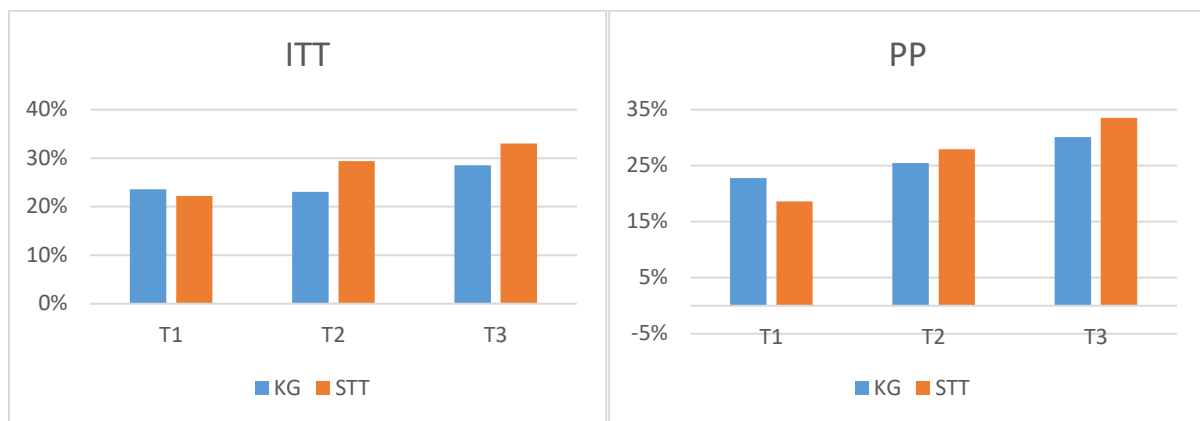


Abbildung 13 Anteil SuS, die angegeben haben, bereits schon einmal geschwänzt zu haben.

Zur Übersicht wurden in **Abbildung 13** die relativen Häufigkeiten der Angabe, dass bereits geschwänzt wurde, dargestellt. Über beide Gruppen steigt die Anzahl der Schüler:innen, welche angeben, bereits geschwänzt zu haben über die Zeit hinweg, wobei der Anstieg in der STT-Gruppe, unerheblich der Unterteilung nach ITT oder PP, steiler ausfällt als in der Kontrollgruppe. Somit liegt der Anteil der Schüler:innen, welche angaben, bereits geschwänzt zu haben, in der STT-Gruppe ab T2 über dem Niveau des Anteils in der Kontrollgruppe. Anhand von χ^2 -Tests wurde auf mögliche Zusammenhänge zwischen Kontroll-, bzw. STT-Gruppe und der Angabe, bereits geschwänzt zu haben geprüft. Es konnten keine signifikanten Zusammenhänge gefunden werden.

Abbildung 14 enthält die relativen Häufigkeiten jeder Antwortkategorie unterteilt nach ITT. Zu T1 und T2 wurden jeweils insgesamt 155 Angaben gezählt. Zu T3 wurden 172 Angaben gezählt. In den Verteilungen fällt auf, dass zu jedem Zeitpunkt die geringste Häufigkeitskategorie „ein Mal“ unter den Schüler:innen der Kontrollgruppe häufiger genannt wurde als unter den Schüler:innen der STT-Gruppe (KG: 48.6%, 46% bzw. 42.1% vs. STT: 30.9%, 33.7% bzw. 31.3%). Hingegen wurden die Antwortkategorie „paar Mal“ in der STT-Gruppe stets häufiger genannt als in der Kontrollgruppe (KG: 28.4%, 39.7% bzw. 41.1% vs. STT: 51.9%, 47.8% bzw. 49%).

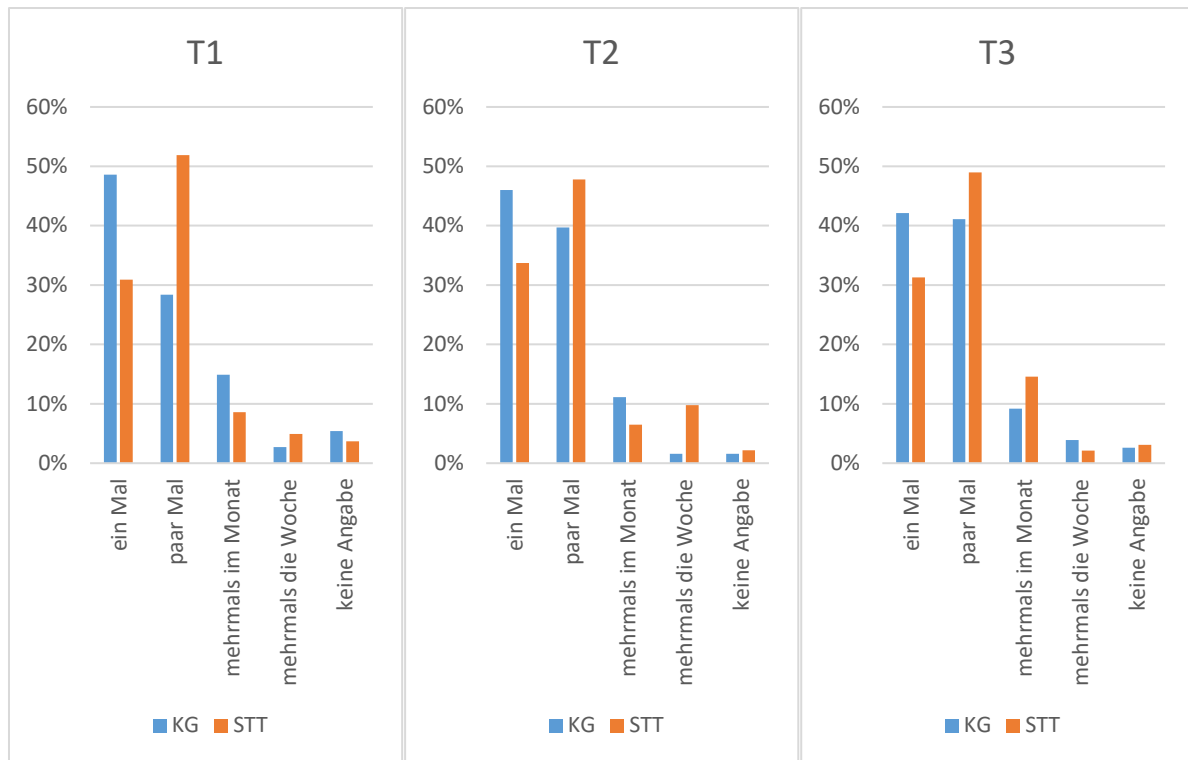


Abbildung 14 Selbstangabe Absentismusverhalten nach ITT.

Kognitives Engagement

Beim kognitiven Engagement zeigte sich sowohl in der ITT-Aufteilung als auch in der PP-Aufteilung ein signifikanter Haupteffekt für Zeit, $F_{ITT}(2, 918) = 9.960, p < .001, \eta_p^2 = .02$ und $F_{PP}(2, 310) = 5.756, p < .01, \eta_p^2 = .02$, und Gruppe, $F_{ITT}(1, 459) = 6.732, p < .05, \eta_p^2 = .01$, $F_{PP}(1, 311) = 6.497, p < .05, \eta_p^2 = .02$. Das kognitive Engagement nahm in allen Schulklassen gleichermaßen über die Zeit ab und unterschied sich nicht zwischen den Gruppen (**Abbildung 15**). Insgesamt berichteten die Schüler:innen der Kontrollgruppe ($n_{ITT} = 217, n_{PP} = 176$) ein signifikant höheres durchschnittliches kognitives Engagement als die Schüler:innen der STT-Gruppe ($n_{ITT} = 244, n_{PP} = 137$).

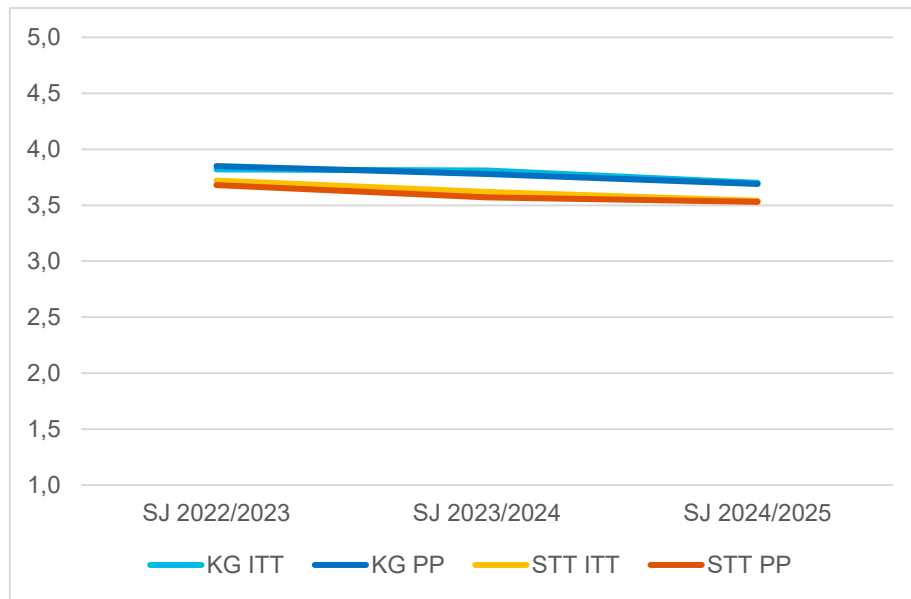


Abbildung 15 Zeitliche Entwicklung des kognitiven Engagements.

Verhaltensbezogenes Disengagement

Sowohl in der Gruppenaufteilung nach PP als auch nach ITT blieb verhaltensbezogenes Disengagement über die Zeit und zwischen Kontroll- und STT-Gruppe stabil auf relativ niedrigem Niveau (siehe **Abbildung 16**). In der inferenzstatistischen Analyse zeigten sich dementsprechend keine signifikanten Unterschiede.

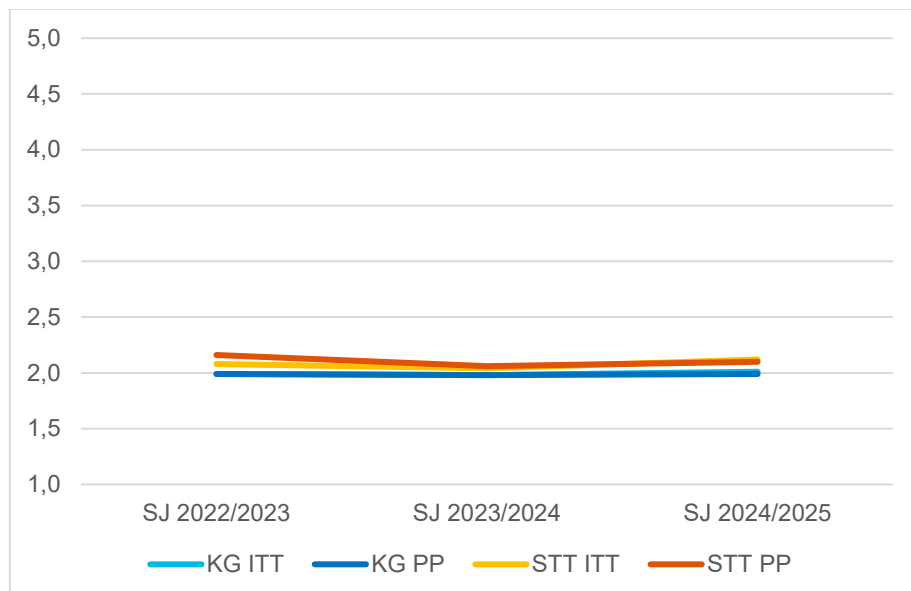


Abbildung 16 Zeitliche Entwicklung des verhaltensbezogenen Disengagements.

Agentisches Engagement

Für agentisches Engagement in der ITT-Gruppenaufteilung konnte kein wesentlicher Unterschied zwischen der STT- und Kontrollgruppe zu T1 festgestellt werden. Während

Schüler:innen in der STT-Gruppe im Zeitverlauf jedoch weniger agentisches Engagement berichteten ($M_{T1} = 2.83$, $M_{T2} = 2.71$, $M_{T3} = 2.69$), zeigte sich in der Kontrollgruppe eine moderate Zunahme ($M_{T1} = 2.81$, $M_{T2} = 2.85$, $M_{T3} = 2.96$). So konnte ein signifikanter Interaktionseffekt zwischen Zeit und Interventionsgruppe festgestellt werden, $F(2, 912) = 6.24$, $p < .01$, $\eta_p^2 = .01$. Demnach unterschieden sich die STT- und Kontrollgruppe im zeitlichen Verlauf des agentischen Engagements signifikant linear, $F(1, 456) = 10.485$, $p < .01$, $\eta_p^2 = .02$.

Bei den Klassen der PP-Gruppenaufteilung ($n_{STT} = 138$, $n_{KG} = 175$) ist in den deskriptiven Werten eine weniger starke Abnahme in der STT-Gruppe ($M_{T1} = 2.75$, $M_{T2} = 2.67$, $M_{T3} = 2.68$) zu beobachten, während sich der Wert in der Kontrollgruppe erst zu T3 nach oben verändert ($M_{T1} = 2.79$, $M_{T2} = 2.80$, $M_{T3} = 2.93$). Hier zeigten sich ebenfalls keine Haupteffekte für Zeit oder Gruppe, und auch der Interaktionseffekt wurde nicht signifikant, d.h. bei STT- vs. Kontrollgruppe liegt das agentische Engagement auf mittlerem Niveau und unterscheidet sich nicht signifikant zwischen den Gruppen.

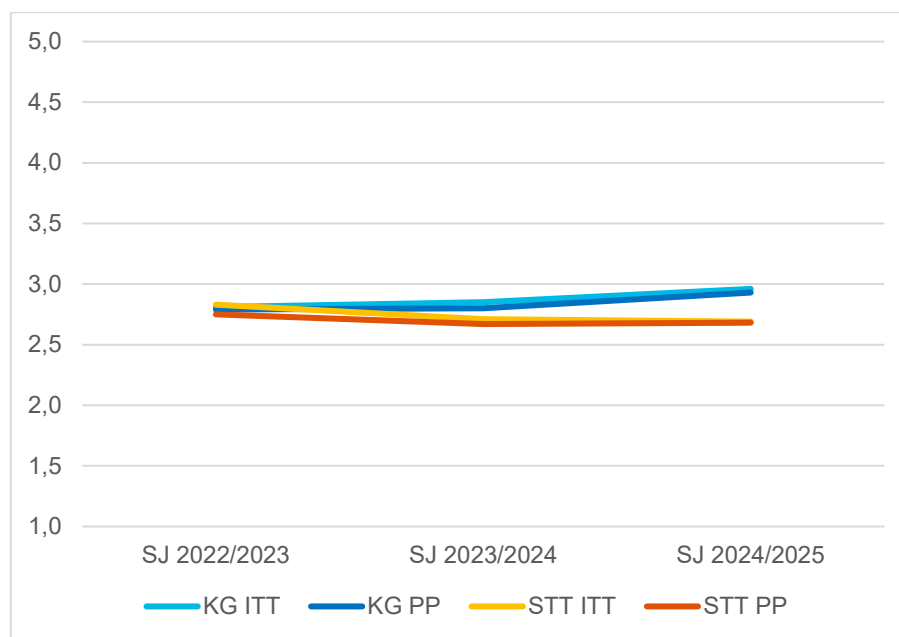


Abbildung 17 Zeitliche Entwicklung des agentischen Engagements.

Motivation: intrinsisch, extrinsisch und identifiziert

Die Mittelwerte und Standardabweichungen für intrinsische, extrinsische und identifizierte Motivation in der STT- und Kontrollgruppe werden separat für die Aufteilung nach ITT und PP in **Tabelle 8** berichtet. In der Aufteilung nach ITT zeigen sich bei der intrinsischen Motivation signifikante Haupteffekte für die Gruppe ($F(1, 459) = 3.28$, $p < .05$, $\eta_p^2 = .01$), die Zeit ($F(2, 918) = 40.31$, $p < .001$, $\eta_p^2 = .08$) sowie deren Interaktion ($F(2, 918) = 4.10$, $p < .05$, $\eta_p^2 = .01$). Die intrinsische Motivation sinkt in allen Klassen über die Zeit hinweg ab ($F(1, 459) = 67.95$, $p < .001$,

$\eta_p^2 = .13$), jedoch stärker in der STT-Gruppe ($F(1, 459) = 6.27, p < .05, \eta_p^2 = .01$). Die Werte für extrinsische Motivation schwanken über die Zeit hinweg nur leicht, weder über die Zeit noch im Vergleich der Gruppen finden sich signifikanten Effekte. Hinsichtlich der identifizierten Motivation findet sich ein Haupteffekt im Gruppenvergleich ($F(1, 459) = 6.79, p < .05, \eta_p^2 = .02$) und ein Haupteffekt der Zeit ($F(1, 918) = 25.53, p < .001, \eta_p^2 = .05$). Unabhängig von der Gruppe nimmt die identifizierte Motivation über die Zeit hinweg ab ($F(1, 459) = 47.69, p < .001, \eta_p^2 = .09$).

In der Aufteilung nach PP und somit in der reduzierten Stichprobengröße bestätigen sich die Effekte im Wesentlichen. Bei der intrinsischen Motivation zeigt sich zwar kein signifikanter Haupteffekt mehr für die Gruppe, die Effekte für die Zeit ($F(2, 564) = 27.89, p < .001, \eta_p^2 = .08$) und die Interaktion ($F(2, 564) = 3.63, p < .05, \eta_p^2 = .01$) bleiben jedoch bestehen. So nimmt der Wert in allen Klassen über die Zeit hinweg ab ($F(1, 282) = 44.37, p < .001, \eta_p^2 = .13$), jedoch stärker in der STT-Gruppe ($F(1, 282) = 4.83, p < .05, \eta_p^2 = .02$). Bei der extrinsischen Motivation zeigen sich keine signifikanten Effekte. Bei der identifizierten Modifikation wird nur der Zeit-Haupteffekt signifikant ($F(2, 564) = 15.69, p < .001, \eta_p^2 = .05$), auch hier ist ersichtlich, wie der Wert in allen Klassen über die Zeit hinweg abnimmt ($F(1, 282) = 28.49, p < .001, \eta_p^2 = .08$).

Tabelle 8 Mittelwerte und Standardabweichungen zu den drei Facetten von Motivation

			T1		T2		T3		
			<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Intrinsisch	ITT	STT	245	3.23	0.89	2.93	0.95	2.76	0.83
		KG	216	3.22	0.93	3.10	0.82	2.97	0.88
	PP	STT	137	3.27	0.89	2.92	0.94	2.79	0.79
		KG	174	3.20	0.94	3.06	0.82	2.95	0.89
Extrinsisch	ITT	STT	245	3.30	0.86	3.30	0.83	3.30	0.77
		KG	216	3.18	0.85	3.23	0.80	3.26	0.74
	PP	STT	137	3.32	0.81	3.30	0.81	3.26	0.81
		KG	174	3.17	0.84	3.23	0.80	3.26	0.75
Identifiziert	ITT	STT	245	3.93	0.81	3.71	0.89	3.60	0.86
		KG	216	4.04	0.80	3.89	0.78	3.78	0.80
	PP	STT	137	3.96	0.75	3.70	0.91	3.59	0.90
		KG	174	3.99	0.80	3.88	0.74	3.80	0.77

Realistische Bildungsaspiration

Im Sinne eines möglichen Schulabbruchrisikos und möglicher Effekte des STT auf ebendieses Risiko wurde in **Tabelle 9** der Anteil jener Schüler:innen festgehalten, welcher angegeben hat, unter den zum Befragungszeitpunkt geltenden Umständen keinen Schulabschluss schaffen zu können. Unabhängig davon, ob nach ITT oder PP differenziert wird, liegt der Anteil der Schüler:innen, der angab, keinen Schulabschluss schaffen zu können ab T2

durchgehend höher in der STT-Gruppe. χ^2 -Tests und Korrelationsanalysen ergaben keine signifikanten Unterschiede. Dies kann unter anderem auf den Umstand zurückzuführen sein, dass die Gesamtzahl der Schüler:innen, welche Angaben, keinen Schulabschluss schaffen zu können, insgesamt relativ gering ist ($n(T1) = 27$; $n(T2) = 28$; $n(T3) = 23$).

Tabelle 9 Anteil Schüler:innen mit realistischer Bildungsaspiration "Kein Schulabschluss"

		T1	T2	T3
ITT	KG	3.87%	2.54%	1.50%
	STT	3.91%	4.27%	4.33%
PP	KG	2.68%	2.20%	1.37%
	STT	3.27%	5.52%	2.91%

2.5.4 Soziales Klassengefüge

Klassenklima

Für das Klassenklima zeigt sich in der Aufteilung nach IPP ein signifikanter Haupteffekt für die Zeit ($F(2, 896) = 11.00, p < .001, \eta_p^2 = .02$) sowie für die Interaktion mit der Gruppe ($F(2, 896) = 9.90, p < .001, \eta_p^2 = .02$). Interessant ist hier der Vergleich über die Zeit hinweg, ersichtlich aus **Abbildung 18**. In der Tendenz nimmt das Klassenklima allgemein über die Zeit ab ($F(1, 448) = 11.34, p < .001, \eta_p^2 = .03$), in der Interaktion zeigt sich jedoch, dass der anfangs höhere Wert in der STT-Gruppe abnimmt, während er in der Kontrollgruppe zu T3 wieder leicht zunimmt und sogar höher liegt als zu T1 ($F(1, 448) = 17.66, p < .001, \eta_p^2 = .04$).

In der Aufteilung nach PP tritt trotz reduzierter Stichprobengröße neben den immer noch signifikanten Effekten (Zeit: $F(2, 610) = 10.08, p < .001, \eta_p^2 = .03$; Interaktion: $F(2, 610) = 7.34, p < .001, \eta_p^2 = .02$) zusätzlich ein signifikanter Haupteffekt der Gruppe auf ($F(1, 305) = 4.92, p < .05, \eta_p^2 = .02$). Der Wert nimmt im Mittel linear über die Zeit ab ($F(1, 305) = 8.65, p < .01, \eta_p^2 = .03$), das Klassenklima „erholt“ sich in der STT-Gruppe zu T3 zwar wie auch in der Kontrollgruppe leicht, liegt jedoch deutlich unter dem Anfangswert ($F(1, 305) = 12.73, p < .001, \eta_p^2 = .04$).

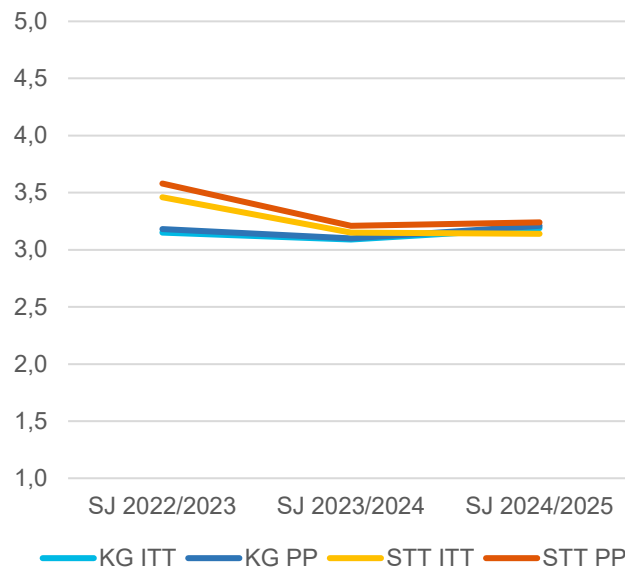


Abbildung 18 Zeitliche Entwicklung des Klassenklimas.

Verhalten innerhalb der Klasse

Die Entwicklung des (negativen) Verhaltens innerhalb der Klasse ist in **Abbildung 19** dargestellt, ein höherer Wert ist hier gleichbedeutend mit mehr wahrgenommenem negativem Verhalten innerhalb der eigenen Klasse.

In der Aufteilung nach ITT zeigen sich bei negativem Verhalten signifikante Haupteffekte für die Gruppe ($F(1, 434) = 10.02, p < .01, \eta_p^2 = .02$), die Zeit ($F(2, 868) = 4.44, p < .05, \eta_p^2 = .01$) sowie deren Interaktion ($F(2, 868) = 7.01, p < .001, \eta_p^2 = .02$). Während das negative Verhalten innerhalb der STT-Gruppe zu T2 zu- und zu T3 wieder abnimmt, geht es in der Kontrollgruppe bereits zu T2 zurück und bleibt zu T3 gleich hoch. Hier zeigt sich also ein linearer Zeiteffekt ($F(1, 434) = 8.30, p < .01, \eta_p^2 = .02$) als auch ein linearer Interaktionseffekt ($F(1, 434) = 5.44, p < .05, \eta_p^2 = .01$).

In der Aufteilung nach PP wird der Haupteffekt der Gruppe nicht mehr signifikant, jedoch aber der Haupteffekt für die Zeit ($F(2, 612) = 3.22, p < .05, \eta_p^2 = .01$) und die Interaktion mit der Gruppe ($F(2, 612) = 4.78, p < .01, \eta_p^2 = .02$). Die Werte für die STT-Gruppe sind in der reduzierten Stichprobe etwas niedriger, auch die Zunahme zu T2 fällt nicht so hoch aus. Es finden sich jedoch erneut ein linearer Zeiteffekt ($F(1, 306) = 5.15, p < .05, \eta_p^2 = .02$) und ein linearer Interaktionseffekt ($F(1, 306) = 6.68, p < .05, \eta_p^2 = .02$).

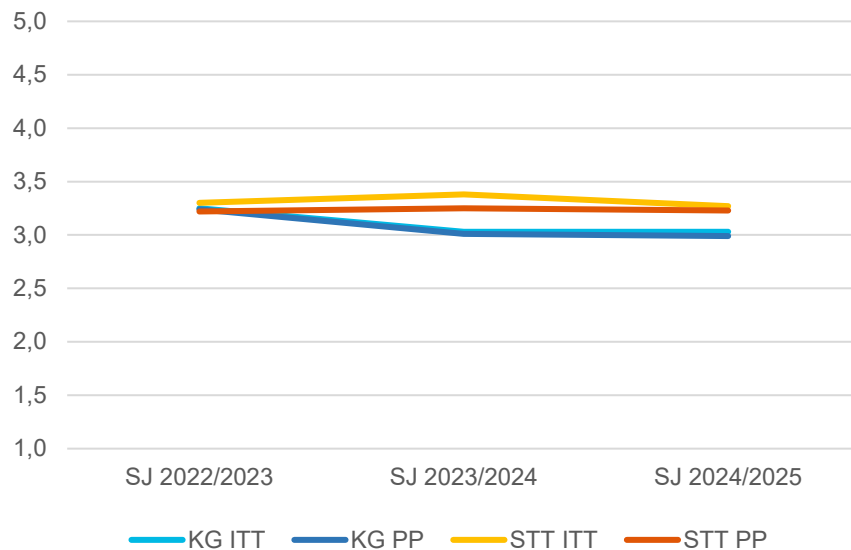


Abbildung 19 Zeitliche Entwicklung des negativen Verhaltens innerhalb der Klasse.

2.6 Zusammenfassung: Ergebnisse der Schüler:innenbefragung

Die Schüler:innen hatten zu drei Messzeitpunkten Gelegenheit, ihre eigene Sicht auf den Schulalltag und ihre Einstellung und Motivation hinsichtlich schulischer Belange zu äußern. Im Vergleich der Kontroll- und STT-Gruppe wird hinsichtlich der demographischen Variablen deutlich, dass diese in Bezug auf das Alter, Geschlecht und sozioökonomischen Hintergrund in ihrer Zusammensetzung vergleichbar sind. Lediglich der Anteil von Schüler:innen mit Migrationserfahrung ist in der Kontrollgruppe signifikant höher.

Zum STT selbst: In Bezug auf die Rolle der Teamteachenden zeigen sich hohe Werte im Vertrauen, allerdings eine signifikante Abnahme des Vertrauens, der wahrgenommenen Fairness sowie der akademischen Unterstützung durch die Teamteachenden über die Zeit hinweg, wobei sich zumindest die akademische Unterstützung zu T3 zu stabilisieren scheint. Im Vergleich zur Kontrollgruppe sinkt auch die wahrgenommene Fairness der Klassenlehrkraft sowie die von ihnen wahrgenommene Autonomieunterstützung über die Zeit stärker, während sie in der Kontrollgruppe über die Zeit stabil bleibt oder leicht zunimmt.

Zu Wohlbefinden und Bedürfniserfüllung: In Bezug auf das Wohlbefinden und die Bedürfniserfüllung ist interessant, dass die Schüler:innen der STT-Gruppe zu T1 teilweise höhere Werte berichten als die der Kontrollgruppe (bspw. beim Wohlbefinden oder beim Erleben von Verbundenheit). Das Kompetenzerleben liegt wiederum durchweg in der Kontrollgruppe höher als in der STT-Gruppe. Während die Erfüllung der Bedürfnisse und das tägliche Wohlbefinden in der STT-Gruppe über die Zeit teilweise recht deutlich abnehmen (bspw. beim Autonomie-Erleben), nehmen die Werte in der Kontrollgruppe leicht zu oder fallen weniger stark ab.

Zum Schulerfolg: In der Literatur wird häufig eine Abnahme von Motivation und Engagement in der Adoleszenz berichtet (bspw. bei Wang & Eccles, 2011). Dieses Muster findet sich auch hier wieder. Auch zeigt sich kein Vorteil der STT- gegenüber der Kontrollgruppe. Beim Absentismusverhalten ist deskriptiv erkennbar, dass der Anteil von Schüler:innen der STT-Gruppe, die angeben zu schwänzen, zwar zu T1 unter dem Anteil der schwänzenden Schüler:innen in der Kontrollgruppe liegt, jedoch über die Zeit stärker zunimmt und zu T2 und T3 deutlich darüber liegt. Das kognitive Engagement liegt bei den Schüler:innen der STT-Gruppe durchweg unter den Werten der Kontrollgruppe, beide nehmen über die Zeit jedoch etwa gleich stark ab. Das verhaltensbezogene Disengagement blieb über die Zeit stabil, hier fanden sich auch keine Gruppenunterschiede. Beim agentischen Disengagement wiederum nahm der Wert in der STT-Gruppe über die Zeit leicht ab, während er in der Kontrollgruppe leicht zunahm. Die intrinsische Motivation nimmt über alle Klassen hinweg über die Zeit ab, jedoch stärker in der STT-Gruppe. Für die extrinsische Motivation lassen sich keine Effekte für Zeit- oder Gruppenunterschiede finden, während bei der identifizierten Motivation wieder beide Gruppen über die Zeit abnehmen, jedoch stärker in der STT-Gruppe.

Zum sozialen Klassengefüge: Ähnlich wie bei der wahrgenommenen Verbundenheit wird in der STT-Gruppe anfänglich höheres Klassenklima berichtet, dieses verschlechtert sich jedoch über die Zeit, während die Schüler:innen der Kontrollgruppe zu T3 ein leicht verbessertes Klassenklima berichten. Negatives Verhalten nimmt in der Klassen mit STT zu T2 zu und geht zu T3 wieder zurück, während es in der Kontrollgruppe bereits zu T2 zurückgeht und sich dann zu stabilisieren scheint.

Zusammenfassend lässt sich festhalten, dass sich über nahezu alle erhobenen Variablen hinweg überwiegend kleine negative Effekte finden lassen, insbesondere was die Interaktion zwischen Zeit und Gruppe angeht, d.h. dass sich über die Zeit für die STT-Gruppe in eine andere Richtung bewegt als die Kontrollgruppe. Vereinzelt finden sich auch mittlere oder größere Effekte (z.B. bei der Abnahme von intrinsischer Motivation über die Zeit hinweg). Es wird deutlich, dass die Ergebnismuster sich kaum zwischen der Aufteilung nach ITT und PP unterscheiden. Daher ist anzunehmen, dass es sich um robuste Effekte handelt, die teilweise auch in der reduzierten Stichprobe bestehen (bspw. beim kognitiven Engagement) oder sogar noch deutlicher hervortreten (bspw. beim Wohlbefinden). In einigen Fällen werden kleinere Effekte in der PP-Bedingung nicht mehr signifikant (bspw. bei der wahrgenommenen Fairness der Lehrkraft) oder verschieben sich leicht (z.B. beim Verhalten innerhalb der Klasse).

TEIL III – ÜBERSICHTEN DER NOTEN UND FEHLTAGE

3.1 Durchführung der Erhebung

Lehrkräfte und Teamteachende wurden per Mail aufgefordert, sog. Übersichten der Noten und Fehltage (ÜNuF) auszufüllen. Im Vorfeld von T2 wurden durch das Erhebungsteam in den ÜNuF die Namen aller Schüler:innen mit vorliegender Einverständniserklärung eingetragen, sodass die Lehrkräfte und Teamteachenden nur noch die Noten und Fehltage eintragen mussten. Diese sensiblen Daten wurden mit hoher forschungsethischer Verantwortung behandelt. Die ÜNuF wurden auf der EAH-hauseigenen Nextcloud passwortverschlüsselt gespeichert. Der Prozess der Datengewinnung und –aufbereitung geschah in Kooperation mit dem Forschungsdatenmanagement HAW-K. Diese Kooperation sicherte die Einhaltung datenschutzrechtlicher Interessen und ermöglichte die Erarbeitung eines wiederholbar konsistenten, python-basierten Einleseverfahrens der ÜNuF. Dieses teil-automatisierte Einleseverfahren ermöglichte, dass die Daten in eine anonymisierte Form zur weiteren Analyse überführt wurden. Auf Empfehlung des Forschungsdatenmanagements wurde zudem für die Aufbereitung der ÜNuF-Daten OpenRefine genutzt.

3.2 Methodik der Datenanalyse

Unverändert zu den vorhergehenden Analysen werden weiterhin Zweifaktorielle Varianzanalysen (ANOVA) mit Messwiederholungen ausgewertet, welche einen Vergleich zwischen Kontroll- und STT-Gruppe über die drei Messzeitpunkte hinweg ermöglichen. In Abschnitt 2.4 ist eine ausführliche Beschreibung der verwendeten Methodik zu finden.

3.3 Ergebnisse

3.3.1 Entschuldigte und unentschuldigte Fehltage

Die Anzahl entschuldigter wie unentschuldigter Fehltage wurde jeweils zum Halb- und Endjahreszeugnis abgefragt. Da die angegebene Anzahl an Fehltagen zum Endjahreszeugnis T1 in manchen Fällen die angegebene Anzahl an Fehltagen zum Halbjahreszeugnis T1 unterbot, wurde geschlussfolgert, dass sich die Angaben für die jeweils betroffene, gesamte Klasse zum Endjahreszeugnis nur auf den Zeitraum des zweiten Halbjahres beziehen. Für diese Fälle wurde die Summe der Fehltage aus den Angaben zum Halb- und Endjahreszeugnis als Wert für das gesamte Schuljahr herangezogen. Um möglicherweise hierin entstehende Fehler zu vermeiden, wurde ab T2 die Frage in die Übersicht der Noten und Fehltage hinzugefügt, ob sich die Angaben der Fehltage zum Endjahreszeugnis auf das gesamte Schuljahr beziehen oder nur auf das zweite Halbjahr. Diese weitere Abfrage ergab indes wiederum neues Fehlerpotenzial, da bspw. angegeben wurde, die Anzahl der Fehltage im Endjahreszeugnis beziehe sich auf das gesamte

Schuljahr, während die zum Endjahreszeugnis angegebene Anzahl deutlich die Anzahl zum Halbjahreszeugnis unterbot. In diesen Fällen wurde die Angabe der Teamteachenden, bzw. Lehrkräften ignoriert und lediglich die Summe beider Angaben als Anzahl der Fehltage für das gesamte Schuljahr erachtet. Es schien, dass würden die Schulen dies unterschiedlich handhaben.

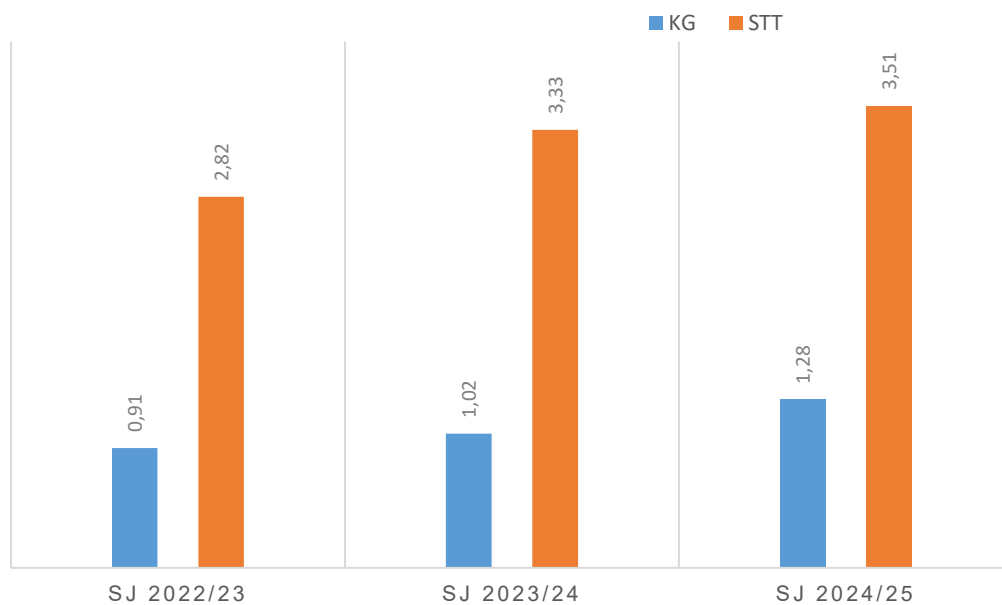


Abbildung 20 Durchschnittliche Anzahl unentschuldigter Fehltage pro Schuljahr pro Schüler:in gruppiert nach ITT.

Wie in **Abbildung 20** dargestellt, liegt die durchschnittliche Anzahl unentschuldigter Fehltage zu T1 in der STT-Gruppe nach ITT um etwa das Dreifache höher als in der Kontrollgruppe ($F(2, 874) = 7.28, p < .01, \eta_p^2 = .016$) und nach PP in den SJ 2022/23 bis 2023/24 sogar um das Vierfache höher ($F(2, 588) = 7.49, p < .01, \eta_p^2 = .025$) (siehe **Tabelle 10**). Es gibt jedoch keinen Anstieg der unentschuldigten Fehltage mit der Zeit in der STT-Gruppe, der signifikant stärker ausfällt als jener in der Kontrollgruppe. Hinsichtlich der entschuldigten Fehltage sind keine signifikanten Differenzen zwischen den beiden Gruppen nach ITT festzustellen (**Abbildung 21**).

Tabelle 10 Unentschuldigte Fehltage

			SJ 2022/23	SJ 2023/24	SJ 2024/25
		<i>n</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
ITT	STT	251	2.82 (8.71)	3.33 (12.5)	3.51 (14)
	KG	188	0.91 (3.2)	1.02 (2.58)	1.28 (3.75)
PP	STT	155	3.46 (10.66)	4.41 (15.62)	4.33 (17.37)
	KG	141	0.58 (2.06)	1.04 (2.55)	1.22 (3.72)

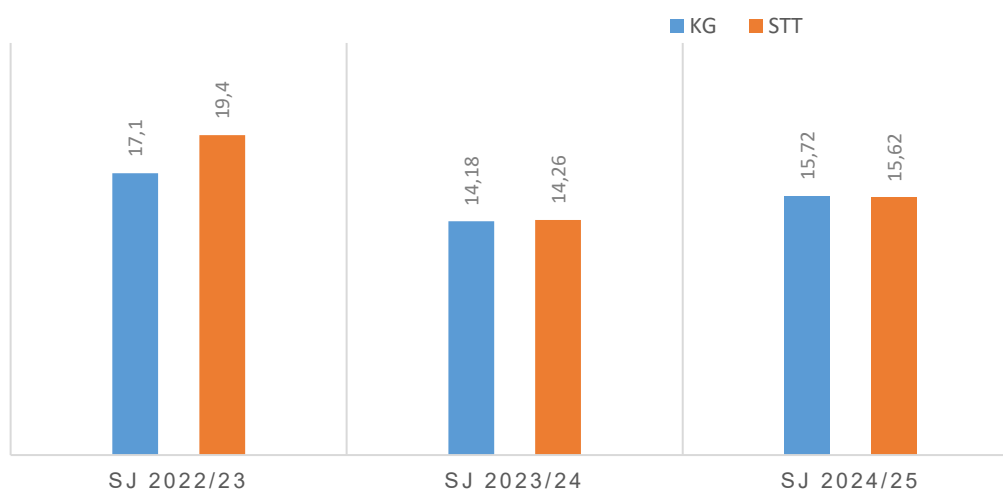


Abbildung 21 Durchschnittliche Anzahl entschuldigter Fehltage pro Schuljahr pro Schüler:in gruppiert nach ITT.

3.3.1 Klassenlistendaten: Schulnoten

Die Übersichten über Noten und Fehltage enthalten Angaben zu den Noten Deutsch, Mathematik und Englisch. Aus diesen wurde ein individueller Notenschnitt berechnet. In **Abbildung 22** ist die zeitliche Entwicklung des Notenschnitts für Kontroll- und STT-Gruppe nach ITT enthalten. Der Notenschnitt verschlechtert sich für beide Gruppen, wobei die Verschlechterung für die STT-Gruppe signifikant mit der Zeit ansteigt, während die Schüler:innen der Kontrollgruppe das zu T2 erreichte Niveau halten (siehe **Tabelle 11**). Die Differenzen zwischen den beiden Gruppen sind schon zu T1 signifikant ($F(2, 1060) = 14.04, p < .001, \eta_p^2 = .026$), wobei die Differenz zwischen den beiden Gruppen mit der Zeit signifikant wächst (Interaktionseffekt Time x Gruppe: $F(3, 529) = 2.67, p < .1, \eta_p^2 = .005$, Test auf linearen Kontrast: $F(2, 530) = 4.27, p < .01, \eta_p^2 = .008$). Während die Differenz zu T1 zwischen den beiden Gruppe noch 0.18 beträgt, ist diese auf 0.29 Notenpunkte angestiegen. Der Notendurchschnitt in der STT-Gruppe liegt zu T3 um 0.3 Notenpunkte höher als zu T1. In der Kontrollgruppe liegt diese Differenz bei 0.19 Notenpunkten.

Tabelle 11 Notenschnitt

			SJ 2022/23	SJ 2023/24	SJ 2024/25
<i>n</i>			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
ITT	STT	309	2.87 (0.76)	3.02 (0.79)	3.17 (0.79)
	KG	223	2.69 (0.69)	2.81 (0.70)	2.88 (0.78)
PP	STT	179	2.90 (0.80)	3.03 (0.83)	3.09 (0.87)
	KG	174	2.70 (0.68)	2.79 (0.70)	2.82 (0.75)

Die Differenzen lassen sich ebenfalls nach Differenzierung nach PP beobachten, wenn auch weniger stark ausgeprägt. Zwar unterscheiden sich die Gruppen unverändert signifikant voneinander ($F(2, 702) = 9.79, p < .01, \eta_p^2 = .027$), die F-Werte des Interaktionseffekts mit der Zeit sowie die Testergebnisse hinsichtlich linearer Kontraste sind nicht mehr signifikant.

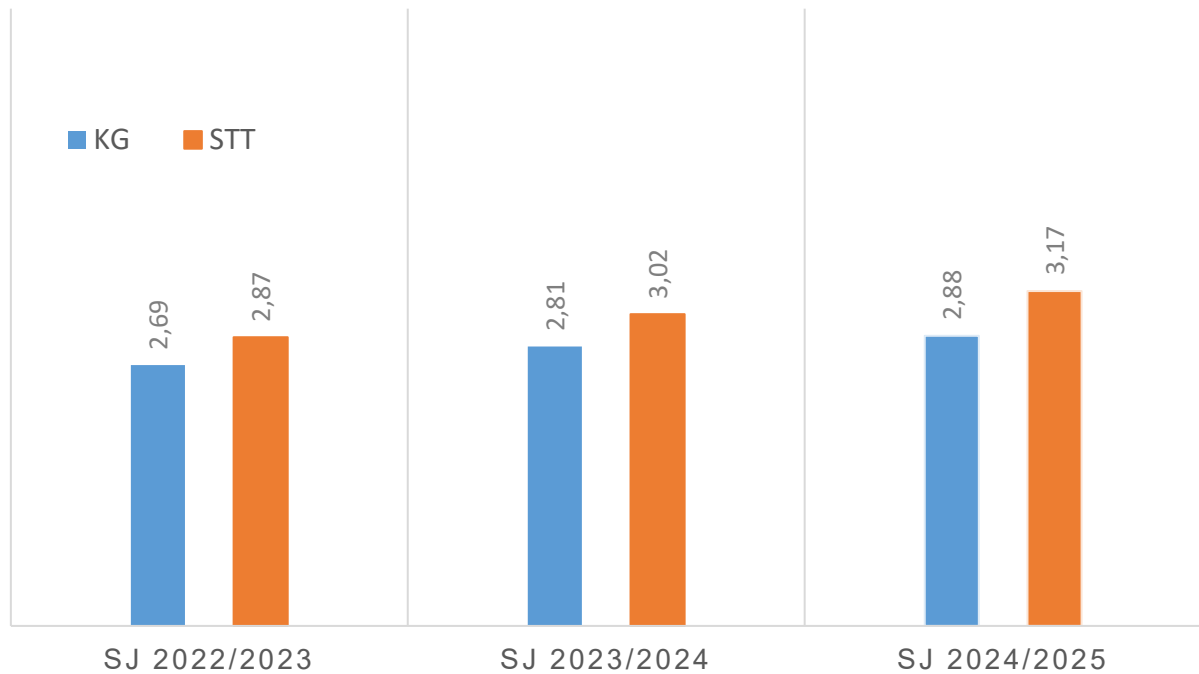


Abbildung 22 Zeitliche Entwicklung des Gesamtnotenschnitts (Deutsch, Mathe und Englisch), Unterteilung nach ITT.

3.4 Zusammenfassung: Ergebnisse der ÜNuF-Erhebung

Die von den Lehrkräften und Teamteachenden ausgefüllten Übersichten über Noten und Fehltage weisen auf starke Differenzen hinsichtlich unentschuldigter Fehltage und Durchschnittsnoten zwischen der STT- und Kontrollgruppe hin. Während bezüglich entschuldigter Fehltage keine signifikanten Differenzen zwischen den Gruppen festzustellen sind, liegt hinsichtlich der unentschuldigten Fehltage bereits zum ersten Messzeitpunkt eine eklatante Diskrepanz zwischen den Gruppen vor, mit einer höheren Anzahl unentschuldigter Fehltage in der STT-Gruppe. Diese Diskrepanz vergrößert sich mit der Zeit jedoch nicht. Auch die Noten stellen sich in der sozialpädagogischen Teamteachinggruppe signifikant schlechter dar als in der Kontrollgruppe. Je nach Gruppierung (ITT vs. PP) stellen sich entweder signifikante Verschlechterungen der STT-Gruppe über die Zeit dar (ITT) oder nicht (PP).

TEIL IV –BEFRAGUNG DER TEAMTEACHENDEN UND KLASSENLEHRKRÄFTE

4.1 Durchführung der Erhebung

Zusätzlich zu den Schüler:innen-Erhebungen fanden Erhebungen mit Teamteachenden und/oder Lehrkräften der Erhebungsklassen mittels Online-Fragebögen statt. Die Erhebungen fanden zeitlich versetzt zu den Befragungen der Schüler:innen statt, jeweils zwischen Oktober und Dezember 2023 (T1) und 2024 (T2). Allen Klassenlehrkräften sowie den Teamteachenden der Erhebungsklassen wurde per Mail ein Link zum Onlinefragebogen zugesendet, innerhalb des Erhebungszeitraums wurden zudem Erinnerungsmails versendet, an der Befragung teilzunehmen. Der Fragebogen erhielt mehrere **Filterfragen** zu Beginn, welche die Einordnung der Rolle (Klassenlehrkraft, Teamteacher:in, Fachlehrkraft) sowie die Zuordnung zu den an der Erhebung beteiligten Klassen ermöglichten.

4.2 Stichprobenbeschreibung

Die Beteiligung war innerhalb der STT-Gruppe etwas höher, die Teilnehmendenzahlen sind **Tabelle 12** zu entnehmen. Von den zu T1 als STT-Klassen erhobenen Klassen nahmen 92% der Teamteachenden teil und 88% der Tandem-Lehrkräfte. In der Kontrollgruppe war die Rücklaufquote mit 68% etwas niedriger. Wie auch in der Analyse der Schüler:innen-Daten wurden die Klassen der ITT und/oder PP-Bedingung zugeordnet, der Unterschied in der Stichprobengröße fiel jedoch geringer aus als in der Schüler:innen-Stichprobe. Insgesamt konnten Daten von 26 Lehrkräften und 14 Teamteachenden analysiert werden, die an beiden Erhebungen teilgenommen hatten.

Tabelle 12 Teilnehmende an der Befragung der Tandems und Klassenlehrkräfte

Erhebungszeitpunkt		STT-Gruppe		Kontrollgruppe	
		Anzahl TT	Anzahl LK	Anzahl TT	Anzahl LK
T1	ITT	23	22	0	17
	PP	13	14	0	12
T2	ITT	13	17	3 ^a	15
	PP	11	11	0	11

Anmerkungen. ^a In einigen der ursprünglich als Kontrollgruppen in die Erhebung aufgenommenen Klassen wurde nach T1 STT eingeführt, daher lagen für drei dieser Klassen zu T2 auch Fragebögen von TT vor.

Der demographische Hintergrund der teilnehmenden Teamteachenden und Klassenlehrkräfte ist **Tabelle 13** zu entnehmen. Alle Angaben basieren auf den zu T1 gemachten Angaben. Auffallend ist die geringe Repräsentanz nicht-deutscher Herkunft sowohl bei den

Teamteachenden als auch den Klassenlehrkräften, lediglich 2 Klassenlehrkräfte gaben zu T1 ein anderes Geburtsland als Deutschland bzw. eine andere Erstsprache als Deutsch zu haben. Insgesamt sind die beteiligten Teamteachenden und Lehrkräfte eher jünger und vorrangig weiblich.

Tabelle 13 Demographischer Hintergrund der teilnehmenden Lehrkräfte und Teamteachenden

		TT		LK	
		absH	relH	absH	relH
Alter	20-29 Jahre	2	4.0	1	2.6
	30-36 Jahre	15	30.0	16	42.1
	40-49 Jahre	5	10.0	7	18.4
	50-59 Jahre	1	2.0	9	23.7
	60-69 Jahre	0	0.0	5	13.2
Geburtsland	Deutschland	23	46.0	36	94.7
	anderes Land	0	0.0	2	5.3
Erstsprache	Deutsch	23	46.0	36	94.7
	andere Sprache	0	0.0	2	5.3
Geschlecht	männlich	3	6.0	5	13.2
	weiblich	20	40.0	33	86.8
Seit wann in der aktuellen	maximal 6 Monaten	8	34.7	4	10.5
	maximal 1 Jahr	3	13.0	1	2.6
Klasse?	maximal 1 Jahr und 6 Monaten	10	43.5	14	36.8
	maximal 2 Jahr	2	8.6	1	2.6
	maximal 2 Jahre und 6 Monate	0	0.0	9	23.7
	3 Jahre und länger	0	0.0	9	23.7

Anmerkungen. absH = absolute Häufigkeit, relH = relative Häufigkeit.

4.3 Verwendete Skalen

Der Fragebogen erhielt neben Fragen zum demographischen Hintergrund (s.o.) Skalen zur Stressbelastung (nur LK), Arbeitszufriedenheit, Identifikation mit der Schule, zum Commitment gegenüber dem STT und Kooperation im Tandem (nur in STT-Gruppe). Diese werden im Folgenden kurz umschrieben.

4.3.1 Stressbelastung

Im Vergleich zu anderen Berufen sind Lehrkräfte zwar weniger physisch gefordert, jedoch mental stark beansprucht. So liegt die Burn-Out-Rate bei 10-30% und es besteht ein hohes Risiko, psychosomatische Beschwerden zu entwickeln (Nübling et al., 2015). Daher ist es wichtig, berufsspezifische Belastungsfaktoren zu identifizieren. Diese Maßnahme wird als Gefährdungsbeurteilung bezeichnet und ist nach dem Arbeitsschutzgesetz (§§ 4-6 ArbSchG) für Arbeitgeber:innen jedes Berufsfelds verpflichtend.

Unsere Erfassung der Belastungsfaktoren bei Lehrkräften basiert auf dem Copenhagen Psychosocial Questionnaire (COPSOQ) von Kristensen et al. (2005), welcher als valides und umfangreiches Messinstrument für psychische Belastungen und Beanspruchungen bei der Arbeit gilt. In der deutschen Standardversion umfasst dieser 30 Skalen mit insgesamt 141 Items sowie separate Dimensionen für Lehrkräfte (Nübling et al., 2015). Für die vorliegende Studie wurde sich an diesen orientiert und neun Items entsprechend re-formuliert. Dafür wurde jede Skala der Dimensionen „Anforderungen“ und „Schulspezifika“ einzeln betrachtet und deren relevantester Inhalt in einem übergreifenden Item zusammengefasst. Diese wurden entsprechend der Zielgruppe für Lehrkräfte sowie Teamteachende formuliert auf einer Skala von 1 (Gar nicht) bis 5 (Sehr) statt wie im Original von 1 (Nie/ fast nie/ in sehr geringem Maß) bis 5 (Immer/ in sehr hohem Maß).

4.3.2 Arbeitszufriedenheit

Die Items der Skala zu **Arbeitszufriedenheit** wurden an die revidierte, erweiterte und deutsche Form des *Job Diagnostic Surveys* von Hackman und Oldham (1975) angelehnt formuliert (Kil et al., 2000). Diese besteht ursprünglich aus sieben Items und erfasst die Bewertung der aktuellen Arbeitstätigkeit. Die Skala von Kil et al. (2000) erzielte eine interne Konsistenz von $\alpha = .80$. In der vorliegenden Studie wurden vier Items verwendet, die auf dem Inhalt zweier Items aufbauen („Allgemein gesprochen bin ich mit meiner Arbeit sehr zufrieden.“ und „Ich bin zufrieden mit meiner Arbeit.“) sowie ein auf den Schulkontext spezifisch formuliertes Item.

4.3.3 Identifikation mit der Schule

Laut Leach et al. (2008) besteht soziale Identifikation aus zwei übergeordneten Dimensionen: *self-definition* und *self-investment*. *Self-definition* meint die wahrgenommene Ähnlichkeit zwischen einem selbst und der Gruppe, *self-investment* hingegen beschreibt die subjektive Wichtigkeit der Gruppenzugehörigkeit, sowie die Stärke der positiven Emotionen gegenüber der Gruppe. Diese Dimensionen sind abhängig von 5 untergeordneten Faktoren: *self-stereotyping*, *in-group homogeneity*, *satisfaction*, *solidarity* und *centrality*. Im Schulkontext

könnte das Gefühl der Zusammengehörigkeit unter dem Kollegium Implikationen für Berufszufriedenheit und Burnoutisiko haben.

In unserer Studie besteht diese Variable aus 3 Fragen, die ursprünglich aus der Sozialen Identitäts-Skala von Roth und Mazziotta (2015) stammen (Deutschsprachig, englisches Original von Leach et al. (2008)). Die entnommenen Fragen stammen aus den Subskalen „Solidarität“ und „Zufriedenheit“. Die Leerstellen der Fragen wurden jeweils so ergänzt, dass die „in-group“ die eigene Schule ist. Zwei der Fragen wurden dabei zur Vereinfachung leicht umformuliert. Aus der Frage „Ich empfinde Solidarität mit [...]“ aus der Subskala „Solidarität“ wurde „Ich setze mich für diese Schule ein.“. Die Frage „Ich bin froh, dass ich [. . .] bin.“ Aus der Subskala „Zufriedenheit“ wurde in „Ich bin froh, dass ich an dieser Schule unterrichte bzw. arbeite.“ umformuliert.

4.3.4 Commitment gegenüber dem STT

Wenn Änderungen in einem System eingeführt werden, ist es wichtig, dass alle Akteur:innen hinter dem neuen Konzept stehen, denn sonst hat jede Änderung schlechte Chancen, erfolgreich zu sein (Herscovitch & Meyer, 2002). Auch für eine erfolgreiche Etablierung des STT ist es wichtig zu wissen, ob das gesamte Kollegium, und insbesondere die Teams selbst, an den Wert dieser Arbeit glauben.

Die Theorie unterscheidet drei verschiedene Arten von Commitment: affektiv, kalkulativ und normativ. Affektives Commitment beschreibt die positiven oder negativen Gefühle des Arbeitnehmers bezüglich seiner/ihrer Institution. Normatives Commitment entspricht dem Grad zu dem sich der:die Arbeitnehmer:in moralisch verpflichtet fühlt, im Job zu bleiben. Kalkulatorisches Commitment hingegen beschreibt den Zustand, wenn Arbeitnehmer:innen aus praktischen Gründen das Gefühl haben, ihren Job jetzt nicht verlassen zu können. Der Gedanke ist, dass verschiedene Commitment-Typen unterschiedlich viel Einsatz in Ihrer Arbeit zeigen, und auch unterschiedlich hohe Burnout-Risiken vorweisen (Webs, 2016).

Wir verwenden Items aus der *Commitment to Organisational Change Scale*, entwickelt von Herscovitch und Meyer (2002). Diese Skala bezieht sich speziell auf die Einstellung der Angestellten zu institutionellen Veränderungen. Wir verwenden allerdings nur die Items aus der Subskala *affective commitment*, welche sich auf die emotionale Einstellung der Arbeitnehmer:innen gegenüber der „Veränderung“ (hier: dem STT), beziehen. Das Alpha für diese Skala in der Originalstudie lag bei .94. Wir haben alle 6 Items der affective commitment Skala übernommen und übersetzt. Das Antwortformat wurde von einer 7-Punkt Skala (strongly disagree – strongly agree) auf eine 5-Punkt Skala verändert, drei Items wurden revers kodiert.

4.3.5 Kooperation im Tandem

Diese Skala bezog sich explizit auf die Zusammenarbeit der Teamteachenden und Klassenlehrkräften im STT-Tandem und wurde daher auch nur von den an der Erhebung

beteiligten STT-Tandems erhoben. Neben dem Commitment zum STT kann auch die Qualität der Tandem-Kooperation als möglicherweise relevant für den Erfolg des STT erachtet werden. Um diese zu erfassen, adaptierten wir aus dem COPSOQ die Dimension pädagogische Vorstellungen, da wir diese als besonders wichtig erachteten. Der COPSOQ wurde ursprünglich in Dänemark entwickelt (Kristensen & Borg, 2000). Es gilt als ein umfassendes und validiertes Instrument zur Erfassung von psychosozialen Belastungen im Berufsalltag, sowie deren Folgen. Der COPSOQ wurde bereits in viele Sprachen übersetzt, wir orientieren uns an der deutschen Standardversion, entwickelt von der Freiburger Forschungsstelle Arbeits- und Sozialmedizin (FFAS; siehe z.B. Nübling et al., 2015), sowie der ergänzten Version des COPSOQs speziell für Lehrkräfte in Haupt- und Realschulen. Die die lehrkräftespezifischen Zusatzaspekte des COPSOQ, in welchen die Dimension pädagogische Vorstellungen enthalten ist, finden sich dabei nur in der Lehrkräfte-Version und wurden vor allem aus dem FASS (Fragebogen zur Arbeitssituation an Schulen; (Kaempff & Krause, 2004) übernommen. Für den Fragebogen wurde der Bezug der Items vom Kollegium bzw. der Arbeit als Lehrkraft auf das Tandem bzw. die Arbeit im STT verändert.

4.4 Methodik der Datenanalyse

Auch für die Angaben der Teamteachende und Lehrkräfte wurden Zweifaktorielle Varianzanalysen (ANOVA) mit Messwiederholungen ausgewertet, welche einen Vergleich zwischen Kontroll- und STT-Gruppe über die drei Messzeitpunkte hinweg ermöglichen. In 2.4 ist eine ausführliche Beschreibung der verwendeten Methodik zu finden. Da hier jedoch nur zwei Messzeitpunkte vorlagen, entsprach der Innersubjektfaktor dem Innersubjektkontrast, sodass nur Haupteffekte für die Zeit und die Gruppe sowie Interaktionseffekte angegeben werden.

4.5 Ergebnisse

Für alle analysierten Variablen sind in **Tabelle 14** Mittelwerte und Standardabweichungen der einzelnen Messzeitpunkte für die Lehrkräfte und Teamteachende in den Erhebungsklassen angegeben, getrennt nach ITT- und PP-Bedingung. Die inferenzstatistische Analyse der einzelnen Skalen ist in den folgenden Abschnitten erläutert.

Tabelle 14 Mittelwerte und Standardabweichungen der TT/LK-Erhebung für T1 und T2

Variable		ITT				PP			
		T1		T2		T1		T2	
		<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
Stressbelastung (LK)	STT	13	3.15 (0.47)	13	3.17 (0.46)	9	3.23 (0.39)	9	3.09 (0.41)
	KG	12	2.52 (0.69)	12	2.53 (0.67)	9	2.57 (0.78)	9	2.53 (0.74)
Stressbelastung (TT)	STT	11	2.64 (0.49)	11	2.62 (0.51)	9	2.60 (0.50)	9	2.54 (0.48)
Arbeitszufriedenheit	STT	14	3.77 (0.51)	14	3.55 (0.97)	9	3.78 (0.52)	9	3.83 (0.85)
	KG	12	4.06 (0.41)	12	4.17 (0.62)	9	4.00 (0.45)	9	4.19 (0.69)

Identifikation mit der	STT	14	3.86 (1.08)	14	4.02 (0.77)	9	3.67 (1.15)	9	4.11 (0.57)
Schule	KG	12	4.42 (0.47)	12	4.44 (0.46)	9	4.37 (0.48)	9	4.52 (0.50)
Commitment (LK)	STT	14	4.74 (0.34)	14	4.89 (0.24)	9	4.70 (0.38)	9	4.98 (0.05)
	KG	12	4.65 (0.62)	12	4.78 (0.25)	9	4.54 (0.69)	9	4.76 (0.25)
Commitment (TT)	STT	11	4.88 (0.28)	11	4.86 (0.31)	9	4.94 (0.17)	9	4.87 (0.33)
Kooperation (LK)	STT	11	4.56 (0.47)	11	4.60 (0.36)	9	4.58 (0.46)	9	4.64 (0.33)
Kooperation (TT)	STT	11	4.31 (0.39)	11	4.25 (0.60)	9	4.40 (0.37)	9	4.24 (0.65)

4.5.1 Stressbelastung

In der Aufteilung nach ITT zeigte sich für die Stressbelastung der Lehrkräfte ein signifikanter Haupteffekt der Gruppe: die Lehrkräfte der STT-Gruppe gaben eine **höhere Stressbelastung** an ($F(1,23) = 9.99, p < .01, \eta_p^2 = .08$). Weder für die Zeit noch für eine Interaktion zwischen Zeit und Gruppe fanden sich signifikante Effekte. In der Aufteilung nach PP trat der Haupteffekt für die Gruppe ebenfalls auf ($F(1,16) = 5.61, p < .05, \eta_p^2 = .26$). Insgesamt liegt die Stressbelastung der Lehrkräfte im unteren bis mittleren Bereich. Dies gilt auch für die Teamteachenden, wobei sich bei diesen ebenfalls kein signifikanter Effekt der Zeit zeigte.

4.5.2 Arbeitszufriedenheit

Die Skala für Arbeitszufriedenheit wurde gemeinsam für Teamteachende und Lehrkräfte in der STT- und Kontrollgruppe betrachtet. Obwohl deskriptiv erkennbar war, dass diese in der Kontrollgruppe höher angegeben wurde, fand sich in der ITT-Aufteilung nur ein marginal signifikanter Haupteffekt für die Gruppe ($F(1,24) = 3.69, p < .10, \eta_p^2 = .13$). In der PP-Aufteilung ließen sich keine signifikanten Effekte zeigen. Die befragten Teamteachenden und Lehrkräfte geben eine mittlere bis hohe Arbeitszufriedenheit an.

4.5.3 Identifikation mit der Schule

Die Identifikation mit der Schule wurde in der STT-Gruppe ebenfalls gemeinsam für Teamteachende und Lehrkräfte betrachtet. Insgesamt ist die Identifikation im mittleren bis hohen Bereich angesiedelt. Deskriptiv betrachtet lagen diese sowohl in der Aufteilung nach ITT als auch nach PP zu T1 weiter auseinander als zu T2, hier berichteten die Tandems der STT-Gruppe eine höhere Identifikation mit der Schule als zu T1. Diese Entwicklung spiegelt sich jedoch nicht in der Interferenzstatistik wider: in der ITT-Bedingung trat lediglich ein marginal signifikanter Haupteffekt für die Gruppe auf ($F(1,24) = 3.79, p < .10, \eta_p^2 = .14$). Für die Veränderung über die Zeit oder eine mögliche Interaktion zeigten sich keine signifikanten Effekte. In der Aufteilung nach PP war der Haupteffekt der Gruppe ebenfalls nur marginal signifikant ($F(1,16) = 3.64, p < .10, \eta_p^2 = .19$), ein Zeit- oder Interaktionseffekt trat ebenfalls nicht auf.

4.5.4 Commitment gegenüber dem STT

Das Commitment gegenüber dem STT wurde getrennt für Lehrkräfte und Teamteachende ausgewertet. Deskriptiv betrachtet war das Commitment in der STT- höher als in der Kontrollgruppe., insgesamt aber auf einem hohen bis sehr hohen Level. Während die Lehrkräfte in der STT-Gruppe zu T1 im Schnitt etwas niedrigeres Commitment berichteten als ihre Teamteachenden-Kolleg:innen, berichteten sie zu T2 im Schnitt höheres Commitment, während die Teamteachenden zu T2 einen im Schnitt niedrigeren Wert berichteten als zu T1. Die Werte sind im Wesentlichen zwischen der Aufteilung nach ITT und nach PP zu vergleichen. Es zeigen sich jedoch in beiden Bedingungen keine signifikanten Haupt- oder Interaktionseffekte. Lediglich bezüglich des Lehrkräfte-Commitments in der PP-Bedingung findet sich ein marginal signifikanter Effekt für die zeitliche Veränderung ($F(1,16) = 3.86, p < .10, \eta_p^2 = .19$).

4.5.5 Kooperation im Tandem

Die Kooperation im Tandem wurde in der STT-Gruppe getrennt für Lehrkräfte und Teamteachende analysiert. Hier weichen die Werte der Lehrkraft minimal ab von denen der Teamteachenden – auch hier zeigt sich ein dem Commitment ähnliches Muster. Während die Kooperation zu T1 durch die Lehrkräfte etwas schlechter bewertet wird als von den Teamteachenden, ist dies zu T2 andersherum. Diese Veränderung über die Zeit spiegelt sich jedoch weder in der Aufteilung nach ITT noch nach PP in einem signifikanten Effekt wider.

4.6 Zusammenfassung: Ergebnisse der LK- und TT-Fragebögen

Bezüglich der Befragung der Lehrkräfte und Teamteachende ist positiv hervorzuheben, wie **hoch die Beteiligung in den STT-, aber auch in der Kontrollgruppe** war. Ungeachtet der Beteiligung handelt es sich aus statistischer Perspektive bei den Stichprobengrößen pro Messzeitpunkt dennoch um „kleine“ Stichproben, insbesondere in der Aufteilung nach PP, sodass **nur ein einziger Haupteffekt (höhere Stressbelastung der Lehrkräfte in der STT-Gruppe) signifikant** wurde. Bezüglich der Befragung der Lehrkräfte und Teamteachenden ist auffällig, wie positiv die Werte ausfallen: insgesamt ist das wahrgenommene Stresslevel noch auf einem verträglichen mittleren Bereich, während eine hohe Arbeitszufriedenheit und Identifikation mit der Schule angegeben werden. Hier spielt ggf. Selbstselektion eine Rolle, dass die Personen auch freiwillig an der Befragung teilnahmen, die ein hohes Commitment zur Schule aufweisen. Unter Berücksichtigung marginal signifikanter Effekte zeigte sich, dass die STT-Tandems niedrigere Arbeitszufriedenheit und eine geringere Identifikation mit der Schule angaben als ihre Kolleg:innen in der Kontrollgruppe.

TEIL V – WEITERFÜHRENDE ANALYSEN

5.1 Einleitung

Aufgrund der in vielen Fällen hypothesenkonträren Ergebnisse sollte in weiterführenden Analysen zunächst untersucht werden, ob eine andere Analysestrategie zu unterschiedlichen Schlussfolgerungen führt. Darauffolgend sollen mögliche Erklärungsansätze für die Ergebnisse gefunden und untersucht werden.

5.2 Analyse der T3-Werte in Abhängigkeit der Dauer an STT

Tabelle 15 enthält die Korrelationskoeffizienten der Gesamtdauer an STT bis zum Zeitpunkt T3 (FB-Daten) bzw. zum Ende des Schuljahres (ÜNuF-Daten, in der Tabelle unter dem Strich zu sehen) mit den von uns untersuchten Ergebnismaßen zum Zeitpunkt T3/Schuljahresende. Es zeigen sich für die Fairness und die Autonomieunterstützung der Klassenlehrkraft, alle Maße zur Bedürfniserfüllung, dem Wohlbefinden im Schulalltag, intrinsische und identifizierte Motivation und realistische Bildungsaspiration signifikant negative und für die idealistische Bildungsaspiration marginal signifikante negative Zusammenhänge mit der STT-Gesamtdauer ($r_s = -.07$ bis $-.19$). Für unterdrückende Emotionsregulation, Disengagement, negatives Verhalten im Klassenverband, Notenschnitt und unentschuldigte Fehltage zeigen sich signifikant positive Zusammenhänge, was bedeutet, dass mit steigender STT-Dauer die Schüler:innen mehr unterdrückende Emotionsregulation, mehr Disengagement, mehr negatives Verhalten im Klassenverband, einen höheren (also schlechteren) Notenschnitt und mehr unentschuldigte Fehltage aufweisen ($r_s = .09$ bis $.17$).

Tabelle 15 Korrelations- und multiple lineare Regressionsanalysen zwischen der Dauer des STT und abhängigen Variablen, kontrolliert auf den ersten Messzeitpunkt

Variable	r T3 mit STT Gesamtdauer	stand. B für AV: T3; UV: STT-Dauer T1 bis T3, kontrolliert für T1
Fairness (LK)	-.13**	-.02
Autonomieunterstützung (LK)	-.19***	-.11**
Bedürfniserfüllung		
Autonomie	-.13**	-.07 ⁺
Kompetenzerleben	-.12**	-.07 ⁺
Verbundenheit	-.14***	-.11**
Wohlbefinden im Schulalltag	-.15***	-.10*
Emotionsregulation		
Neubewertung	.00	-.05
Unterdrückung	.10*	.02
Schulabsentismus	.04	.04
Engagement		
kognitiv	-.08 ⁺	-.03
agentisch	-.05	-.04

Disengagement		.09*	.01
Motivation	intrinsisch	-.09*	-.04
	identifiziert	-.11**	-.09*
	extrinsisch	.02	-.05
Klassenklima		-.02	-.07
Negatives Verhalten im Klassenverband		.14***	.08*
Bildungsaspiration	idealistisch	-.07*	-.05
	realistisch	-.09*	-.07*
Notenschnitt (Hauptfächer)		.17***	.06*
Anzahl der Fehltag	unentschuldigt	.14***	-.02
	entschuldigt	.07	.03

Anmerkungen. UV = unabhängige Variable, AV = abhängige Variable. Gesamtdauer STT berechnet sich aus der Summe der einzelnen Tage, in denen einer Klasse ein:e Teamteacher:in zugeteilt war.
 *** $p < .001$. ** $p < .01$. * $p < .05$. + $p < .10$

Um zu prüfen, ob die STT-Dauer zwischen T1 und T3 mit den Ergebnismessungen am Ende des Schuljahres zusammenhängt, nachdem das **Ausgangsniveau** der jeweiligen Variablen berücksichtigt wurde, wurde außerdem für jedes Ergebnismaß eine multiple lineare Regressionsanalyse gerechnet. In diesen Regressionsanalysen fungierten die Werte der Ergebnismaße zum Zeitpunkt T3/letztes Schuljahr als abhängige Variable und die STT-Dauer zwischen T1 und T3 sowie die Werten der Ergebnismaße zum Zeitpunkt T1 als unabhängige Variablen. Die zweite Spalte in **Tabelle 15** enthält die standardisierten Regressionskoeffizienten der STT-Dauer zwischen T1 und T3 aus diesen Analysen.

In diesen Analysen zeigten sich für Autonomieunterstützung der Lehrkraft, Bedürfniserfüllung Verbundenheit, Wohlbefinden im Schulalltag und identifizierte Motivation signifikant negative Effekte und für Bedürfniserfüllung Autonomie und Kompetenzerleben sowie realistische Bildungsaspiration marginal signifikante negative Effekte der STT-Dauer zwischen T1 und T3 (β s = -.07 bis -.11). Für negatives Verhalten im Klassenverband zeigten sich zudem signifikant positive und für Notenschnitt marginal signifikant positive Regressionskoeffizienten (β s = .06 bis .08), was darauf hindeutet, dass mehr STT-Zeit mit mehr negativem Verhalten und einem schlechteren Notenschnitt einhergeht. Auf Wunsch der Kooperationspartner wurden außerdem Analysen gerechnet, bei denen nur jene 6 Klassen als STT-Klassen eingingen, die über 3 Jahre hinweg ohne jegliche Unterbrechung der Tandemzusammenarbeit STT stattfand. Hierbei sollte beachtet werden, dass diese Analysen auf einem stark idealisierten Szenario basiert, das der tatsächlichen Umsetzungspraxis des STT an den Schulen nicht gerecht wird. Sie bezieht sich lediglich auf eine sehr spezifische Teilgruppe – 6 von 25 Klassen, in denen über drei Jahre hinweg weder Ausfälle noch Personalwechsel auftraten – und lässt sich daher realistischerweise nicht flächendeckend übertragen. Vergleicht man die T3-Ergebnismaße zwischen diesen 6 Klassen mit STT mit der Kontrollgruppe der PP-Differenzierung ergeben sich für diese Analysen zunächst

deutlich kleinere Stichproben ($N = 298$), was die Möglichkeit beschränkt signifikante Effekte zu finden. Dennoch zeigen sich auch hier für Emotionsunterdrückung als Emotionsregulationsstrategie ($M_{STT} = 3.39$ ($SD_{STT} = 0.80$), $M_{KG} = 3.09$ ($SD_{KG} = 0.85$), $t(285) = -2.63$, $p < .01$), realistische Bildungsaspiration ($M_{STT} = 2.78$ ($SD_{STT} = 0.74$), $M_{KG} = 2.99$ ($SD_{KG} = 0.73$), $t(289) = 2.11$, $p < .05$), Notenschnitt ($M_{STT} = 3.19$ ($SD_{STT} = 0.90$), $M_{KG} = 2.87$ ($SD_{KG} = 0.80$), $t(296) = -2.95$, $p < .01$) und unentschuldigte Fehltage ($M_{STT} = 4.91$ ($SD_{STT} = 21.90$), $M_{KG} = 1.45$ ($SD_{KG} = 4.97$), $t(284) = -2.12$, $p < .05$) signifikant schlechtere Ergebnisse in der STT-Gruppe zum Messzeitpunkt T3.

5.3 Erklärungsansätze

Für die durchgängig signifikant hypothesenkonträren oder nichtsignifikanten Zusammenhänge der Ergebnismaße mit dem STT kommen aus unserer Sicht verschiedene Erklärungsansätze in Betracht, die wir im Folgenden näher untersuchen wollten.

5.3.1 Statistische Artefakte & Stichprobencharakteristika

Zunächst stellten wir uns die Frage, ob z.B. durch verstärkte Rekrutierungsbemühungen (Teilnahme an der Befragung) der Teamteachenden in der STT-Gruppe vermehrt auch Schüler:innen mit größerer Schuldistanz teilnehmen, die in der Kontrollgruppe ebenfalls vertreten sind, jedoch aufgrund fehlender Einverständniserklärungen nicht teilnahmen. Die Rücklaufrate war in der Tat positiv korreliert mit der Gesamtdauer an STT ($r = .26^{***}$) und signifikant höher in der STT-Gruppe (ITT: $M = 73\%$ vs. 62% , $t(755) = -11.11$, $p < .001$; PP: $M = 73\%$ vs. 62% , $t(525) = -8.98$, $p < .001$). Jedoch korrelierte die Rücklaufrate lediglich mit dem Kompetenzerleben ($r = -.13$), dem alltäglichen Wohlbefinden ($r = -.10$), dem agentischen Engagement ($r = -.10$), der intrinsischen Motivation ($r = -.08$), der realistischen Bildungsaspiration ($r = -.14$) und dem Notenschnitt ($r = .11$) signifikant. Bei Kontrolle des Zusammenhangs der STT-Dauer mit diesen Variablen für die Rücklaufrate, war dieser für Kompetenzerleben ($\beta = -.10$), das alltägliche Wohlbefinden ($\beta = -.15$), sowie die intrinsische Motivation ($\beta = -.09$) unverändert negativ und für den Notenschnitt ($\beta = .15$) unverändert positiv signifikant, während lediglich der Zusammenhang mit realistischer Bildungsaspiration nicht länger signifikant war. Demnach scheiden verstärkte Rekrutierungsbemühungen der Teamteachenden in der STT-Gruppe für die meisten Ergebnismaße als Erklärung aus.

Eine weitere potenziell denkbare Erklärung wäre es, wenn Schüler:innen mit schlechten Noten in der STT-Gruppe eher versetzt werden würden als in der Kontrollgruppe, wodurch Schüler:innen mit „schlechteren“ Werten in der Kontrollgruppe eher zum Schuljahreswechsel die evaluierte Klasse verlassen würden. Jedoch zeigte sich kein signifikanter Unterschied in der Versetzungswahrscheinlichkeit zwischen der STT- und Kontrollgruppe (ITT: 97.8% vs. 97.8% , $\chi^2(1, N = 681) = 0.03$, $p = .96$; PP: 98.3% vs. 97.8% , $\chi^2(1, N = 470) = 0.16$, $p = .69$) zum Ende des

Schuljahres 2022/23. Am Ende des Schuljahres 2023/24 zeigten sich für die ITT-Einteilung sogar etwas niedrigere Versetzungsraten in der STT-Gruppe im Vergleich zur Kontrollgruppe (ITT: 95.1% vs. 98.1%, $\chi^2(1, N = 687) = 4.72, p < .05$; PP: 95.4% vs. 97.8%, $\chi^2(1, N = 465) = 1.97, p = .16$). Kontrolliert man diesen Zusammenhang jedoch für die Versetzungsgefahr anhand der Schulnoten 5 und 6, so gibt es auch hier keinen Zusammenhang zwischen der Gruppeneinteilung und den Versetzungsraten ($\chi^2(1) = 1.53, p = .22$).

5.3.2 Selektionshypothese

In den ÜNuF fragten wir klassenweise ab, ob die Klassen für das STT speziell zusammengestellt wurden (z.B. nach Förderschwerpunkt) und erfragten zusätzlich bei den Schulleitungen telefonisch, ob es vor der Einführung des STT Unterschiede zwischen den Klassen gab und anhand dieser Klassen speziell für das STT ausgewählt wurden. Im Folgenden untersuchten wir, inwieweit dies der Fall war und welchen Einfluss dies ggf. auf die Ergebnisse hatte.

Zunächst schlossen wir alle Klassen, für die in den ÜNuF zu T1, T2 oder T3 eine spezielle Zusammensetzung angegeben wurde oder die eine parallele Kontrollgruppen zu einer entsprechend zusammengesetzten STT-Klasse waren, aus dem Datensatz aus und führten mit der erhaltenen **Substichprobe erneut Analysen** durch ($n = 504$). Für 12 der 14 zuvor in der Gesamtstichprobe signifikanten Zusammenhänge zwischen der STT-Gesamtdauer und den Ergebnismaßen änderte sich durch die Beschränkung auf diese Subgruppe weder etwas an der Signifikanz noch an der Richtung der Zusammenhänge. Lediglich für unterdrückende Emotionsregulation ($\beta = .08^*$) und realistische Bildungsaspiration ($\beta = -.09^*$) zeigte sich nicht länger ein signifikanter, sondern nur noch ein marginal signifikanter Zusammenhang. Das kognitive Engagement hingegen, welches in der Gesamtstichprobe nur marginal signifikant mit der STT-Dauer assoziiert war, hatte in der Substichprobe ohne spezifisch zusammengesetzte Klassen sogar einen signifikant negativen Zusammenhang mit der STT-Dauer ($\beta = -.09^*$).

In telefonischen Interviews konnten weiterhin 14 der 16 Schulleitungen erreicht werden. Für 4 Klassen konnte den Angaben der Schulleitung zufolge eine gezielte Zuordnung des STT sicher ausgeschlossen werden. Für 19 der 50 Klassen (STT und KG) konnte hingegen sicher von einer Zuordnung des STT zu diesen Klassen innerhalb der Schule ausgegangen werden (siehe zur Zusammensetzung von Schulklassen S.13). Bei **Ausschluss** dieser 19 Klassen, denen gezielt STT zugewiesen wurde und ihren entsprechenden Parallelklassen, sofern sie an der Erhebung in Form von Kontrollgruppen teilnahmen, zeigte sich für die verbleibenden Klassen bei 12 der 14 zuvor in der Gesamtstichprobe signifikanten Zusammenhänge zwischen der STT-Dauer und den Ergebnismaßen **keine Änderung**, weder in der Signifikanz noch in der Richtung der Zusammenhänge. Lediglich für die Bedürfniserfüllung nach Verbundenheit ($\beta = -.07$) und

negatives Verhalten des Klassenverbands ($\beta = .05$) waren die zuvor in der Gesamtstichprobe signifikanten Zusammenhänge mit der STT-Dauer in der Substichprobe ohne gezielte Zuordnung zu STT- oder KG-Bedingung nicht länger signifikant. Allerdings gab es in der Substichprobe signifikant negative Zusammenhänge von kognitivem Engagement ($\beta = -.13^*$) sowie idealistische Bildungsaspiration ($\beta = -.17^{**}$) und signifikant positive Zusammenhänge von Schulabsentismus ($\beta = .14^*$) mit der STT-Dauer, die in der Gesamtstichprobe noch nicht vorlagen. Demnach zeigen sich auch in nicht spezifisch zusammengesetzten Klassen und nicht gezielt den STT- oder KG-Bedingungen zugeordneten Klassen ähnliche (viele) signifikante Zusammenhänge schlechterer Werte mit längerer STT-Dauer, weswegen die hier beobachteten Effekte und Unterschiede nicht durch eine vorab stattgefundene Selektion der Schüler:innen oder Klassen zurückzuführen sind.

5.3.3 Stigmatisierungshypothese

Die Stigmatisierungshypothese fußt auf der Idee des Pygmalion- bzw. Rosenthaleffekts. Dieser beschreibt, wie Erwartungen an eine Person deren Leistung positiv oder negativ beeinflusst (Rosenthal, 2002). Hieraus lässt sich die Hypothese bilden, dass mit höherer Erwartung der Lehrkräfte an ihre Schüler:innen, diese tatsächlich höhere Motivation, Engagement und bessere Schulnoten zeigen oder vice versa (niedrige Erwartung – wenig Anstrengung). Weiterhin konnte die Hypothese aufgestellt werden, dass sich die Erwartungen der Lehrkräfte an die eigenen Schüler:innen zwischen der STT- und Kontrollgruppe unterscheiden, d.h. wir prüften, ob Schüler:innen der STT-Gruppe mit niedrigeren Erwartungen konfrontiert und sich dies in ihren tatsächlichen Leistungen widerspiegeln würde.

Die Erwartungen der Lehrkräfte an ihre Schüler:innen wurden mittels dreier Variablen operationalisiert: Erwartung über Schulabschluss, Annahme über herausfordernde familiären Bedingungen, Erwartung hinsichtlich des Entwicklungspotenzials der Schüler:innen). Hinsichtlich des Anteils der Schüler:innen, denen ein **Abschluss zugetraut** wird, gibt es keine signifikanten Gruppenunterschiede. Weiterhin wurde die Lehrkraft-Erwartung anhand der Auffassung, wie hoch der Anteil der Schüler:innen ist, welcher aus herausfordernden familiären Verhältnissen kommt. Hier wird das oft bestätigte Axiom zugrunde gelegt, dass die Leistungserwartung mit der Einschätzung der familiären, sozialen und ökonomischen Verhältnisse der Schüler:innen zusammenhängt (Gentrup et al., 2018). Zum ersten Befragungszeitpunkt der Lehrkräfte gaben diese signifikant höhere Werte in der STT-Gruppe an ($F = 107.56, p < .001$). **Abbildung 23** enthält die Werte entsprechend der ITT-Aufteilung. Für die PP-Aufteilung ergaben sich zum ersten Messzeitpunkt noch stärkere Diskrepanzen (KG: 43.06% vs. STT: 72.19%, $F = 170.104, p < .001$). Zu T2 hingegen gehen die Differenzen zwischen den beiden Gruppen zurück (ITT), bzw. kehren sich um (PP: KG: 61.5% vs. STT: 55.48%, $F = 6.929, p < .001$).

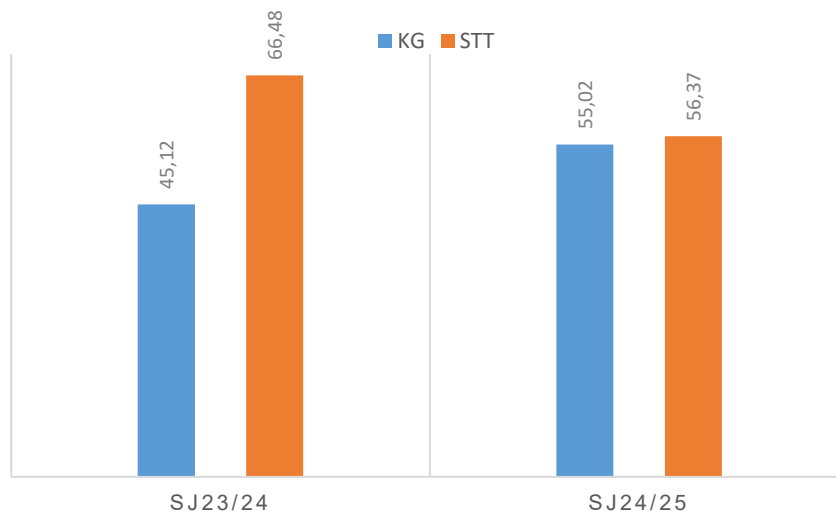


Abbildung 23 Anteil SuS, die nach Auffassung der LK aus herausfordernden familiären Verhältnissen kommt, Angaben in Prozent, Differenzierung nach ITT.

Zuletzt wurde die Erwartungshaltung der Lehrkraft anhand ihrer Einschätzung operationalisiert, ob die Schüler:innen alles in allem **eine gute Zukunftsperspektive** haben oder nicht. Nach ITT-Differenzierung wurde zu beiden Erhebungszeitpunkten eine deutlich positivere Zukunftsperspektive für die Schüler:innen der Kontrollgruppe gesehen (T1: $F = 82.805$, $p < .001$, T2: $F = 44.304$, $p < .001$, siehe **Abbildung 24**). Die Differenzen nach PP sind vergleichbar, wobei zu T2 deutlichere Diskrepanzen zwischen den Gruppen festzustellen ist als nach ITT (T1: KG: 8.138 vs. STT: 6.658, $F = 55.354$, $p < .001$; T2: KG: 7.853 vs. STT: 5.899, $F = 60.411$, $p < .001$).

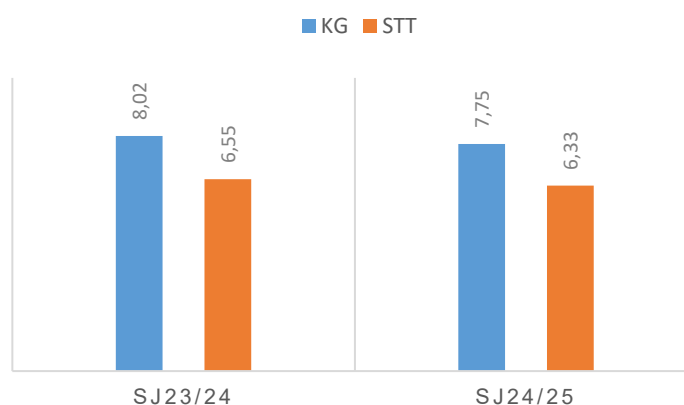


Abbildung 24 Zustimmung der LK zur Aussage, dass die SuS eine gute Perspektive haben, Differenzierung nach ITT.

Um zu prüfen, inwiefern diese unterschiedlichen Erwartungen die Effekte des STT auf die gemessenen Variablen vermitteln, d.h. erklären können, wurden Mediationsanalysen gerechnet,

welche in in **Tabelle 16**, **Tabelle 17** und **Tabelle 18** abgebildet sind. Während Modell 1 keine Mediation enthält, ist in Modell 2 die Milieu-Einschätzung bzw. in Modell 3 die Perspektiven-Einschätzung als Mediator aufgenommen. In den Tabellen angegeben sind die jeweiligen Beta-Koeffizienten der Effektstärken.

Tabelle 16 Ergebnisse der Mediationsanalysen hinsichtlich des Notenschnittes zu T3

	Modell 1	Modell 2	Modell 3
Effekt des STT nach PP-Aufteilung	0.292 ***	0.144 +	0.234 *
Milieu-Einschätzung		0.005 **	
Perspektiven-Einschätzung			-0.036 +
Effekt des STT nach ITT-Aufteilung	0.324 ***	-0.242***	0.282 ***
Milieu-Einschätzung		0.004 **	
Perspektiven-Einschätzung			n.s.

Anmerkungen. n.s. = nicht signifikant. *** $p < .001$. ** $p < .01$. * $p < .05$. + $p < .10$.

Tabelle 17 Ergebnisse der Mediationsanalysen hinsichtlich der intrinsischen Motivation zu T3

	Modell 1	Modell 2	Modell 3
Effekt des STT nach PP-Aufteilung	-0.158 +	n.s.	-0.161 +
Milieu-Einschätzung		n.s.	
Perspektiven-Einschätzung			n.s.
Effekt des STT nach ITT-Aufteilung	-0.194 **	-0.196 **	-0.195 **
Milieu-Einschätzung		n.s.	
Perspektiven-Einschätzung			n.s.

Anmerkungen. n.s. = nicht signifikant. *** $p < .001$. ** $p < .01$. * $p < .05$. + $p < .10$.

Tabelle 18 Ergebnisse der Mediationsanalysen hinsichtlich des agentischen Engagements zu T3

	Modell 1	Modell 2	Modell 3
Effekt des STT nach PP-Aufteilung	-0.192 *	-0.194 +	n.s.
Milieu-Einschätzung		n.s.	
Perspektiven-Einschätzung			n.s.
Effekt des STT nach ITT-Aufteilung	-0.230 **	-0.243 **	0.225 *
Milieu-Einschätzung		n.s.	
Perspektiven-Einschätzung			n.s.

Anmerkungen. n.s. = nicht signifikant. *** $p < .001$. ** $p < .01$. * $p < .05$. + $p < .10$.

Wie aus diesen Tabellen ersichtlich, können die Differenzen in Milieueinschätzung und Einschätzung der Zukunftsperspektive die negativen Effekte des STT, unabhängig ob ITT oder PP, auf agentisches Engagement und intrinsischer Motivation nicht mediierten, da sie selbst keine

signifikanten Effekte auf die Ergebnismaße zeigten. Ergebnisse in Bezug auf andere Operationalisierungen von Motivation und Engagement wurden hier aufgrund ebenfalls ausbleibender Effekte nicht dargestellt. Weiterhin konnten keine Mediationseffekte der Einschätzung der Zukunftsperspektive auf den Effekt des STT auf den Notenschnitt festgestellt werden (**Tabelle 16**). Hingegen wurden Mediationseffekte der Milieueinschätzung auf den Notenschnitt gefunden. So konnten die Effekte des STT sowohl für ITT als auch PP zumindest partiell durch die Aufnahme der Milieueinschätzung moderiert werden.

Um das Verhältnis zwischen Milieueinschätzung, Notengebung und STT weiter aufzuklären, wurden darüber hinaus zeitvariante Mehrebenenanalysen gerechnet. In diesen Modellen konnte jeweils ein signifikanter Effekt des STT als auch der Milieueinschätzung durch die Lehrkraft auf die Notengebung festgestellt werden. Allerdings führen diese Ergebnisse eher zu der Interpretation, dass die Milieueinschätzung einen eigenständigen Effekt darstellt, der unabhängig vom STT die Notengebung beeinflusst.

Insgesamt lässt die Befundlage keine klaren Schlüsse darauf zu, dass die negativen Effekte des STT durch die Lehrkraftmilieueinschätzung auf die Notengebung mediiert, bzw. erklärt werden können. Vor dem Hintergrund, dass sich die STT- und Kontrollgruppe hinsichtlich des sozioökonomischen Status, den wir abgefragt hatten, nicht statistisch signifikant unterscheiden, sind die unterschiedlichen Einschätzungen und Erwartungen der Lehrkräfte hinsichtlich des familiären Hintergrunds eher als **Projektion** des Lehrpersonals zu verstehen. Die Rolle des STT bei der Entstehung dieser Projektionen ist nicht geklärt. Es ist aber denkbar, dass die Lehrkräfte insbesondere gegenüber Schüler:innen der STT-Gruppe (unbewusste) Vorannahmen hinsichtlich deren Leistungsvermögen entwickeln und sich dies in „selbsterfüllenden Prophezeiungen“ zumindest teilweise realisiert. Während die Befundlage hinsichtlich des Effekts dieser Projektion auf die Notengebung nicht eindeutig ist, können in Bezug auf Motivation und Engagement keine negativen Effekte festgestellt werden.

5.3.4 Autonomiehypothese

Ein weiterer Ansatz zur Erklärung der negativen Effekte des STT auf eine Anzahl von Ergebnismaßen ist die Autonomiehypothese. In ihr wird der Annahme nachgegangen, dass die durchgehende Präsenz von zwei Personen in Klassen mit STT, verglichen zu einer Lehr-Person in Klassen ohne STT, die Bedürfniserfüllung nach Autonomie bei den Schüler:innen einschränkt. Wie in Abschnitt 2.5.2 bereits dargelegt, ist zu beobachten, dass zwischen den beiden Gruppen das Erleben von Autonomie zu T1 auf einem vergleichbaren Niveau liegt, bis sie über den Zeitverlauf für Schüler:innen in der STT-Gruppe signifikant unter dem Niveau der Kontrollgruppe sinkt. Eine ähnliche Entwicklung ist bezüglich der wahrgenommenen Autonomieunterstützung durch die Lehrkraft zu beobachten (siehe Abschnitt 2.5.1): Auch hier starten beide Gruppen auf

einem vergleichbaren Niveau. Während der Mittelwert für die Kontrollgruppe zu T2 leicht absinkt und zu T3 sogar wieder steigt, sinkt die wahrgenommene Autonomieunterstützung in der STT-Gruppe konstant über den Zeitverlauf auf ein signifikant niedrigeres Niveau ab.

Nach der Self-Determination-Theorie (SDT) von Ryan und Deci (Ryan & Deci, 2000) ist Erfüllung der Grundbedürfnisse Autonomie, Kompetenzerleben und soziale Eingebundenheit essentiell zur Ausbildung von Motivation. Motivation hingegen gilt als Grundbaustein, auf dem schulischer Erfolg basiert (siehe auch Abschnitt 2.2.1). Demnach ist von der Nichterfüllung des Autonomiebedürfnisses eine Reihe an negativen Effekten auf eine Vielzahl an Ergebnismaßen zu erwarten.

Evaluationsrelevant ist hierbei die Frage, ob und inwiefern ein geringeres Autonomieerleben dazu beiträgt, die negativen Effekte des STT zu erklären. Hierzu wurde eine Reihe an Mediationsanalysen durchgeführt, deren Ergebnisse in **Tabelle 18** dargestellt sind. Die Spalte links gibt die Zielvariable wieder, die drei Spalten rechts davon den Effekt des STT nach verschiedener Operationalisierung auf ebendiese Zielvariable. „*De facto*“ entspricht hier einer bisher nicht verwendeten Operationalisierung. Hier wird eine Klasse der STT-Gruppe zugeordnet, wenn sie mindestens 91 Tage STT erhalten hat. Die drei Spalten rechts des mittleren Teilstriches geben die Effekte des STT wieder, nachdem die Erfüllung des Autonomiebedürfnisses in das Modell mit aufgenommen wurde. Sofern Autonomiebedürfnis selbst einen direkten Effekt auf die Zielvariable hat und der Effekt des STT seine Signifikanz nach Aufnahme des Autonomiebedürfnisses verliert, kann von einer vollständigen Mediation durch die Autonomiebedürfniserfüllung gesprochen werden. Nimmt hingegen nur die Effektstärke des STT ab, bzw. bleibt auf einem niedrigeren Niveau signifikant, so deutet sich eine partielle Mediation durch die Autonomiebedürfniserfüllung ab.

Tatsächlich lassen sich die stärksten Mediationseffekte der Autonomiebedürfniserfüllung auf die Zielvariablen intrinsische Motivation, identifizierende Motivation und kognitives Engagement beobachten. Abgesehen vom STT-Effekt nach „de-facto“-Operationalisierung auf die identifizierende Motivation können alle STT-Effekte **vollständig durch Unterschiede in der wahrgenommenen Autonomie** zwischen den Gruppen erklärt werden. Auch die negativen STT-Effekte auf eine Reihe von weiteren Variablen, wie Kompetenzerleben, Bedürfnis nach sozialer Verbundenheit, tägliches Wohlbefinden, störendes Verhalten in der Klasse und Emotionsregulation konnten zumindest partiell durch die Diskrepanzen in der Autonomiebedürfniserfüllung zwischen den Gruppen erklärt werden. Allerdings konnte kein Effekt der Autonomiebedürfniserfüllung auf Schulnoten gefunden werden. Autonomiebedürfniserfüllung ist also nicht in der Lage, den negativen Effekt von STT auf die Notengebung zu erklären.

Tabelle 19 Mediationsanalysen zur Autonomiebedürfniserfüllung

	Effekt			Effekt nach Mediation			Mediation ITT/PP/d.f.
	ITT	PP	de facto	ITT	PP	de facto	
Kompetenz- erleben	-0.18**	-0.21**	-0.17**	-0.14**	-0.15*	-0.12 ⁺	partiell/partiell/ partiell
Bedürfnis: Verbundenheit	-0.23**	-0.28**	-0.27**	n.s.	-0.18*	-0.19*	vollständig/ partiell/partiell
Tägl. Wohlbefinden	-0.59**	-0.93***	-0.70**	-0.40*	-0.67**	-0.48*	partiell/partiell/ partiell
Intrinsische Motivation	-0.19**	-0.16 ⁺	-0.19**	n.s.	n.s.	n.s.	vollständig/ vollständig/ vollständig
Identifizierende Motivation	-0.17*	-0.20*	0.24**	n.s.	n.s.	-0.12 ⁺	vollständig/ vollständig/ partiell
Verhalten in der Klasse	0.25***	0.27**	-0.15*	0.22**	0.23*	0.26***	partiell/partiell/ partiell
Emotionsregul ation: Suppression	0.16*	0.19*	-0.17*	n.s.	0.14 ⁺	0.16*	vollständig/ partiell/partiell
Notenschnitt	0.32***	0.29***	0.29***	0.31***	+0.23*	0.33***	keine
Kognitives Engagement	-0.16*	-0.16 ⁺	0.19**	n.s.	n.s.	n.s.	vollständig/ vollständig/ vollständig

Anmerkungen. n.s. = nicht signifikant. *** $p < .001$. ** $p < .01$. * $p < .05$. ⁺ $p < .10$.

5.4 Zusammenfassung: Weiterführende Analysen

In diesem Abschnitt wurden mehrere Ansätze geprüft, die zum Ziel haben, die negativen Effekte des STT zu erklären. In Abschnitt 5.3.1 konnte ausgeschlossen werden, dass bestimmte Stichprobencharakteristika und/oder statistische Artefakte für die negativen Effekte des STT verantwortlich sind. Auch scheinen Hintergrundvariablen, die mit einer spezifischen Selektion von entweder Klassen oder Schüler:innen in die STT-Gruppe einhergehen, nicht die negativen Effekte des STT erklären zu können (Abschnitt 5.3.2). Mediationsanalysen zeigen, dass zumindest Teile des negativen STT-Effekts auf die Schulnoten durch eine geringere Leistungserwartung der Lehrkräfte zu erklären ist. Zeitvariante Mehrebenenanalysen deuten jedoch an, dass die geringere Leistungserwartung der Lehrkräfte einen eigenständigen Effekt darstellt, der parallel zum STT die Notengebung der STT-Gruppe negativ beeinflusst (Abschnitt 5.3.3). Weitere Mediationsanalysen ergaben, dass die geringere Autonomiebedürfniserfüllung in der STT-Gruppe eine Vielzahl an negativen Effekten des STT mindestens partiell erklären kann. So können die negativen Effekte des STT auf Motivation und Engagement vollständig erklärt, die negativen Effekte auf

Kompetenzerleben und tägliches Wohlbefinden teilweise erklärt werden. Die Einschränkung des Autonomieerlebens durch die Präsenz von zwei Personen im Klassenraum stellt bisher den stärksten Erklärungsfaktor der negativen Wirkung des STT dar.

TEIL VI – SCHLUSS

6.1 Zusammenfassung

Der vorliegende Projektbericht **schließt das Forschungsprojekt „Evaluation des Sozialpädagogischen Teamteachings an Thüringer Schulen: eine Studie zur Erforschung der Wirk- und Gelingens-Faktoren (EvaTeamTeach)“** ab, das **zwischen Februar 2023 und Dezember 2025** an der Ernst-Abbe-Hochschule Jena durchgeführt wurde. Dafür wurden an **16 Thüringer Schulen** Daten von Schüler:innen, Teamteachenden und Klassenleitungen in insgesamt **50 Schulklassen** ausgewertet. Über drei (Schüler:innen) bzw. zwei (LK und TT) Messzeitpunkte hinweg konnten so Daten von **882 Schüler:innen, 25 Teamteachenden und 44 Klassenlehrkräften** erfasst werden.¹³ Aufgrund der engagierten Mitarbeit und Unterstützung der KSB, der Schulleitungen, Teamteachenden und Klassenlehrkräften an den beteiligten Schulen, waren die Rücklaufquoten im Vergleich zu anderen Evaluationsstudien sehr zufriedenstellend (siehe dazu Abschnitt 2.1.3). Die Ergebnisse der einzelnen Befragungsstandteile wurden bereits in den Abschnitten 2.6, 3.3 und 4.6 zusammengefasst. Mögliche Erklärungsansätze wurden im Abschnitt 5.4 **Fehler! Verweisquelle konnte nicht gefunden werden..** zusammengefasst. Im Folgenden werden die Erhebungsbestandteile überblickshaft zusammengeführt.

Im Rahmen des Projektberichts zur vergangenen Förderperiode war bereits auf Ebene der Schüler:innen eine negative Entwicklung zentraler Variablen über die Zeit (bspw. Bedürfniserfüllung, Engagement und Motivation) in den Klassen mit STT erkennbar (Mitte et al., 2023). **Diese Negativentwicklung lässt sich mit den vorliegenden Analysen bestätigen. Für 11 (ITT, PP: 10) der 22 Ergebnismaße der Fragebogenerhebung findet sich ein signifikant negativer Verlauf über die Zeit,¹⁴ für 10 Ergebnismaße finden sich Nulleffekte (ITT, PP: 11). Der einzige „positive“ Effekt findet sich beim negativen Verhalten innerhalb des Klassenverbunds – dieses nimmt über die Zeit in der STT-Gruppe zu.** Im Vergleich zur Kontrollgruppe zeigen sich mehrfach Interaktionseffekte, bspw. sinkt die Autonomie-Bedürfniserfüllung in der STT-Gruppe über die Zeit stärker ab als in der Kontrollgruppe. Bei 7 Ergebnismaßen (ITT, PP: 4) finden sich durchgängig höhere Werte in der Kontrollgruppe, lediglich das negative Verhalten innerhalb der Klasse ist in der STT-Gruppe stärker verbreitet (nach PP wird der Haupteffekt aber nicht mehr signifikant). Hinsichtlich der Noten und Fehltage **wurden für die**

¹³ Die hier berichteten Zahlen beziehen sich auf die Anzahl derjenigen, die an mindestens einer Erhebung teilgenommen haben. Die Stichprobengrößen für die einzelnen MZP sowie für die Analyse über alle MZP hinweg sind den Abschnitten 2.1.2 und 4.2 zu entnehmen.

¹⁴ Einzige Ausnahme bildet dabei eine „Positiv“-Entwicklung des negativen Verhaltens im Klassenraum – dieses nimmt über die Zeit zu.

Schüler:innen der STT-Gruppe signifikant mehr unentschuldigte Fehltage protokolliert, die Differenz zur Kontrollgruppe veränderte sich über die Zeit nicht. Auch die **Noten der Fächer Deutsch, Mathematik und Englisch stellen sich in der STT-Gruppe signifikant schlechter dar als in der Kontrollgruppe**. Je nach Gruppierung (ITT vs. PP) stellen sich entweder signifikante Verschlechterungen der STT-Gruppe über die Zeit dar (ITT) oder nicht (PP). In der **Befragung der Teamteachenden und Lehrkräfte** zeigte sich lediglich ein signifikanter Effekt: die **Stressbelastung der Lehrkräfte war höher in der STT-Gruppe**. Insgesamt wurden jedoch hohe Werte bspw. bei der Arbeitszufriedenheit oder der Identifikation mit der Schule berichtet.

Wie bereits im Abschnitt 1.2 zur relevanten internationalen Forschungsliteratur herausgearbeitet wurde, erscheinen diese Null- und Negativeffekte nicht unerwartet. So konnte festgestellt werden, dass sich die **Befunde zwischen den Studien deutlich unterscheiden, und von einer generellen Effektivität des Teamteachings nicht ausgegangen werden kann** (siehe dazu bspw. Weerdt et al., 2024). Am ehesten lassen sich positive Effekte auf Ebene der Lehrpersonen finden (siehe dazu Vangrieken et al., 2015). Auf Ebene der Schüler:innen finden sich neben einzelnen positiven Effekten (bspw. bei Ebokaiwe et al., 2024; Vangrieken et al., 2015; Vembye et al., 2024) überwiegend Nullergebnisse (z.B. Johnson et al., 1995; Jurkowski & Müller, 2018; Stefanidis et al., 2023) oder negative Effekte (bspw. Zigmond & Baker, 1990). **In Bezug auf das sozialpädagogische Teamteaching liefern der Abschlussbericht der vergangenen Förderperiode (Mitte et al., 2023) sowie der vorliegende Abschlussbericht die ersten, wissenschaftlich fundierten und belastbaren Ergebnisse.**

6.2 Limitationen und offene Fragen

Wie bereits in Abschnitt 2.4 erwähnt, ist bei der Interpretation der Ergebnisse zusätzlich zu beachten, dass nur jene Akteur:innen und Variablen in die Evaluation einbezogen werden können, die tatsächlich erhoben wurden. Daher sollte die **Interpretation der Ergebnisse stets im Kontext der erhobenen Daten** erfolgen. Beispielsweise liegen keine Daten von Lehrkräften vor, die einzelne Schulfächer in den Klassen unterrichten, und damit im Schulalltag ebenso ein „Tandem“ mit den Teamteachenden bilden. Ein Fokuswechsel in den STT-Manualen von der individuellen Entwicklung der Schüler:innen hin zur Schulentwicklung (siehe dazu im Vergleich Kindersprachbrücke Jena e. V., 2019, 2024) konnte in der Evaluationskonzeption nicht mehr berücksichtigt werden. Allerdings sind Effekte im Sinne dieses Fokuswechsels auch nicht durch Ergebnisse der ersten Förderperiode (Mitte et al., 2023) untermauert. Ob und in welchem Sinne das STT auf Schulebene wirken kann, konnte und kann mit den vorliegenden Daten nicht wissenschaftlich fundiert beantwortet werden.

Eine weitere Limitation stellt die **Varianz in der Ausgestaltung des STT vor Ort** dar. Wie bereits eingangs erwähnt (siehe dazu Abschnitt 1.1) und im Rahmen der vorherigen Förderperiode

erfasst (Mitte et al., 2023), ist das Interventionskonzept des STT stark praxisorientiert. Welche Aufgaben die Teamteachenden im Schulalltag übernehmen und wie sich dadurch die Aufgabenfelder der Klassen- und Fachlehrkräfte verändern, variiert nicht nur von Schule zu Schule, sondern auch von Klasse zu Klasse. Die Teamteachenden sind untereinander gut vernetzt und auch die Klassenlehrkräfte zunehmend eingebunden in Gesprächsformate, was die Weitergabe von best practice Beispielen u.ä. erleichtert. Ob und wann jedoch das STT beginnt, wie lange es dauert, welche Aufgaben die Teamteachenden übernehmen dürfen/können/müssen – dazu gibt es keine bindenden Maßgaben für die Förderung des STT an den Schulen.

Wie bereits in Abschnitt 1.3 ausgeführt, kam es im Verlauf der Evaluation trotz erfolgter Rückversicherungen vonseiten des Evaluationsteams sowie klarer Auswahlkriterien zu zahlreichen STT-Wechseln und Abbrüchen. Zudem gab es eine **große Varianz hinsichtlich der Dauer der Intervention in den STT-Klassen zu T1**. Obwohl diese Varianz bei der Implementation von Teamteaching-Maßnahmen auch in anderen Studien verbreitet zu sein scheint (siehe dazu die Meta-Synthesen von Weerdt et al., 2024 und Vembye et al., 2024), ist die Vergleichbarkeit der Implementation innerhalb der Interventionsgruppe eine der wichtigen Voraussetzungen für eine erfolgreiche Evaluation (Gollwitzer & Jäger, 2014). Wir weisen darauf hin, dass es sich hier um eine Feldstudie handelt und nicht um ein vollständig randomisiertes Design; dennoch wurden wie oben berichtet für viele bekannte Einflussflussgrößen kontrolliert. Im Rahmen unserer Evaluation wurden mehrere Maßnahmen ergriffen, um die Belastbarkeit der Ergebnisse zu prüfen und mögliche Einflüsse der STT-Dauer auszuschließen, bspw. die Analyse nach per-protocol- und intention-to-treat-Ansatz. Dass diese verschiedenen Herangehensweisen an die Auswahl der Klassen als STT- oder Kontrollgruppe dennoch nahezu deckungsgleiche Ergebnisse produzieren, lässt vermuten, dass die negativen oder Null-Effekte der Varianz in STT-Dauer und - Unterbrechungen standhalten und als belastbar aufzufassen sind.

Neben den bereits genannten Limitationen ist zudem auf zentrale Erkenntnisse der **Implementationsforschung** hinzuweisen. Zahlreiche Interventionsstudien zeigen, dass selbst wissenschaftlich gut begründete Programme dann keine Wirkung entfalten, wenn die Implementationsbedingungen vor Ort nicht ausreichend gewährleistet sind (Durlak & DuPre, 2008). Wirkung setzt voraus, dass ein Programm konsistent, vollständig und über einen hinreichenden Zeitraum umgesetzt wird („fidelity“) und dass dokumentiert wird, was tatsächlich durchgeführt wurde und mit welchen Methoden. Wie in Abschnitt 1.1 und 1.3 ausgeführt, waren zentrale Implementationsmerkmale des STT, z.B. die Dauer, Kontinuität, Zuständigkeiten usw. im Feld sehr heterogen. Unter solchen Bedingungen ist es strukturell schwierig, die intendierten Wirkmechanismen eines Programms zuverlässig auszulösen oder empirisch nachzuweisen. Aus Sicht der Implementationsforschung wäre daher zu erwarten, dass selbst ein potenziell

wirksames Programm in dieser Form nur eingeschränkt Wirkung entfalten kann. Dieser Befund entlastet die Programmatik des STT insofern, als die vorliegenden Null- oder Negativbefunde nicht zwingend als Hinweis auf grundsätzliche Unwirksamkeit zu interpretieren sind, sondern vielmehr auf erhebliche Implementationsbarrieren hinweisen, die einer systematischen Wirkung entgegenstehen können.

Ergänzend ist ein weiterer Aspekt aus der klassischen Evaluationslogik nach Cook und Campbell (1979) zu berücksichtigen: sogenannte ausgleichende Rivalität (*compensatory rivalry*). In schulischen Feldstudien kommt es häufig vor, dass Lehrkräfte in Kontrollklassen, bewusst oder unbewusst, motivierter unterrichten, um zu zeigen, dass ihre Klasse ohne Intervention ebenfalls gute Leistungen erbringen kann. Dies kann zu verstärktem Engagement oder großzügigerer Benotung führen und somit Effekte zugunsten der Kontrollgruppe erzeugen. Da weder Unterrichtsintensität noch Leistungsbeurteilungen der Fach- und Klassenlehrkräfte systematisch erfasst wurden, kann ein solcher Rivalitätseffekt im Rahmen dieser Evaluation nicht ausgeschlossen werden. Auch dieser Aspekt könnte – falls er aufgetreten ist – einen Teil der Befundlage erklären und relativiert die Interpretation der negativen oder fehlenden Effekte des STT zuungunsten der Interventionsgruppe.

6.3 Empfehlungen

Ausgehend von der internationalen Forschungsliteratur, den berichteten Ergebnissen, sowie der Prüfung möglicher Erklärungsansätze sollen im Folgenden Empfehlungen gegeben werden, wie und in welche Form das STT zukünftig entwickelt, bewertet und angewendet werden könnte.

6.3.1 Empfehlung 1: Stärkerer Fokus auf Autonomie

Wie im Abschnitt 5.3.4 herausgearbeitet wurde, scheint die Bedürfniserfüllung für Autonomie einen starken Einfluss auf andere zentrale Variablen zu haben und konnte Unterschiede über die Zeit und zwischen den Gruppen vollständig (bspw. intrinsische und identifizierte Motivation) oder partiell (z.B. tägliches Wohlbefinden) aufklären. Daher nehmen wir an, dass die ständige Präsenz zweier (Lehr-)Personen einen signifikant negativen Einfluss auf das Autonomie-Bedürfnis der Schüler:innen hat. So mag das erhöhte Monitoring auf den ersten Blick hilfreich erscheinen (z.B. ob Hausaufgaben eingetragen wurden), aber auf lange Sicht zu einer Wahrnehmung von Autonomieverlust bei den Schüler:innen führen. Über die Zeit berichten die Schüler:innen jedoch keinen Vertrauensverlust gegenüber den Teamteachenden. **Wir empfehlen daher, dass die offenbar als vertrauensvoll erlebte Beziehung zu den Teamteachenden einen Rahmen bietet, um den erlebten Autonomieverlust zu**

kompensieren. So sollten Selbstkompetenzen gefördert werden, die gleichzeitig die Eigenständigkeit der Schüler:innen fördern, bspw. selbstverantwortliche Vorbereitung auf den Unterricht oder der Erwerb von Lernkompetenzen.

6.3.2 Empfehlung 2: Intensität und Dauer des STT zielgruppengerechter gestalten

Ebenfalls negativ beeinflusst sein könnte die Autonomie-Erfüllung durch die erlebte (Un-)Freiwilligkeit des STT. So berichten bspw. Moser Opitz et al. (2021), dass die erlebte Freiwilligkeit einer Teamteaching-Maßnahme positive Auswirkungen haben kann. Da es jedoch keine einheitliche Vorgabe dazu gibt, wann und für wie lange das STT in einer Klasse implementiert werden soll, kann die erlebte Freiwilligkeit bei den Schüler:innen sehr variieren – insbesondere, wenn das STT ihrer Klasse „zugeordnet“ wird (siehe dazu auch Abschnitt 5.3.2). Eine weitere Option wäre es, das STT auf konkrete Fächer zu beschränken, insbesondere zu naturwissenschaftlichen Fächern liegen hier (allerdings ausschließlich zur Zusammenarbeit zweier Lehrkräfte) positive Befunde vor (z.B. Ebokaiwe et al., 2024; Rexroat-Frazier & Chamberlin, 2019). **Wir empfehlen daher, das STT zumindest innerhalb der Schulen nicht nach wechselnden Kriterien einzelnen Klassen zuzuordnen, sondern bspw. für einen festen Zeitraum einer Jahrgangsstufe zuzuordnen oder einem konkreten Schulfach. Dies könnte insbesondere in jüngeren Klassenstufen oder in Übergangsphasen hilfreich sein.**

6.3.3 Empfehlung 3: Rollenklarheit fördern

Wie bereits in Abschnitt 6.2 erwähnt, sind in dieser Evaluation keine Daten von Fachlehrkräften enthalten. Wenn Klassenlehrkräfte und Teamteachende auch abseits der gemeinsam unterrichteten Stunden sehr viel enger zusammenarbeiten und idealerweise ausreichend Zeit zur Verfügung steht für gemeinsame Absprachen (siehe dazu Mitte et al., 2023), ist der Schulalltag doch wesentlich geprägt von der Zusammenarbeit der Teamteachenden mit anderen Lehrkräften. Wie diese Zusammenarbeit aussieht, **könnte eine nicht unerhebliche Auswirkung auf die Wirkung des STT** haben, insbesondere was die Benotung der Schüler:innen in anderen Fächern oder die Bewertung des Disziplinarverhaltens betrifft. Auch die Verortung der Teamteachenden innerhalb der Schule war ein Thema, das bereits in der vergangenen Förderperiode relevant war (siehe dazu Teil 4 in Mitte et al., 2023). Positiv ist hervorzuheben, wie hoch die Tandems bspw. die Identifikation mit der Schule, das Commitment zum STT und ihre Kooperation im Tandem bewerten. Wie wir bei den Erhebungen vor Ort beobachten konnten (siehe dazu auch Abschnitt 6.2), gestalten die Tandems ihre Zusammenarbeit jedoch sehr verschieden. **Wir empfehlen daher, nicht nur den STT-Tandems, sondern auch dem gesamten schulischen Personal konkrete Handlungsrahmen vorzulegen, welche Rolle und Aufgabenfelder die Teamteachenden in ihrer Klasse und an der jeweiligen Schule übernehmen können/sollen – und welche genau nicht.**

6.3.4 Empfehlung 4: Reflexion der Wirkebene

Das Ziel des STT auf Individual- und Klassenebene nach dem Manual von 2019 lautete „alle Schüler:innen der Klasse in ihrer Entwicklung zu eigenverantwortlichen, gemeinschaftsfähigen Persönlichkeiten zu unterstützen und ihnen einen erfolgreichen Schulbesuch zu ermöglichen.“ (Kindersprachbrücke Jena e. V., 2019, S. 9). Unsere Ergebnisse stützen diese positive Entwicklung in einer Reihe von zielrelevanten Variablen jedoch nicht. Während das STT auch und mittlerweile verstärkt auf weitere Wirkungsebenen im Schulsystem eingeht (z. B. Sorgeberechtigte; Kindersprachbrücke Jena e. V., 2019, 2024), fokussierte unsere Untersuchung auf die Zielgruppe der Schüler:innen. Dabei wurde das STT so evaluiert, wie es in den drei Jahren tatsächlich umgesetzt wurde, einschließlich struktureller Herausforderungen wie Personalmangel, längeren Krankheitsausfällen und häufigen Wechseln der Teamteachenden.

Vor diesem Hintergrund stellt sich die Frage, wie das STT zukünftig ausgerichtet sein sollte und welche Voraussetzungen erfüllt sein müssen, damit es seine Wirkungen auf Schüler:innen-Ebene tatsächlich entfalten kann. **Daher empfehlen wir erstens, die Ziele des STT deutlich klarer und wissenschaftlich fundierter zu definieren sowie zu berücksichtigen, dass Ziele verschiedener Wirkungsebenen sich gegenseitig beeinträchtigen könnten.** Unklare oder zu breit gefasste Ziele könnten zu Rollenunsicherheiten, widersprüchlichen Erwartungen und Handlungen führen. Zudem muss transparent sein, welche Effekte realistisch durch das STT erreicht werden können und wo Grenzen bestehen, z. B. im Bereich Schulentwicklung.

Zweitens sollte transparent sein, ob der Schwerpunkt auf der individuellen Förderung sozio-emotionaler Kompetenzen liegt oder ob eine breitere Veränderung schulischer Haltungen und Strukturen angestrebt wird. Beide Zielrichtungen sind legitim, bringen jedoch sehr unterschiedliche Anforderungen mit sich. Ohne diese Zielklarheit besteht das Risiko, dass individuelle sozio-emotionale Entwicklungen, z.B. eine verbesserte Wahrnehmung von Bedürfnissen und Emotionen seitens der Schüler:innen, auf ein Umfeld treffen, das diese Veränderungen nicht aufgreift. Dies kann zu einer Diskrepanz zwischen den Erwartungen der Schüler:innen und der Reaktion bzw. Haltung der Lehrkräfte, Schulleitung oder Eltern führen und dadurch zu verstärkter Schulentfremdung führen (für eine Übersicht siehe Morinaj & Hascher, 2019).

6.3.5 Empfehlung 5: Vulnerabilität der Schüler:innen im Fokus

An Thüringer Regel-, Gesamt- und Gemeinschaftsschulen lässt sich ein großer Bedarf beobachten, sozialer Benachteiligung entgegenzuwirken. Die Abbruchquote an diesen Schulen ist vom bereits hohen Wert von 10.33% im Schuljahr 2017/18 weiter auf 11.35% im Schuljahr 2023/24 gestiegen (Holl et al., accepted for publication). Dabei korreliert die Abbruchquote stark mit dem sozioökonomischen Status (ebd.). Maßnahmen und Projekte wie das STT, die

beabsichtigen, insbesondere dieser sozialen Benachteiligung im Bildungssystem entgegenzuwirken (Kindersprachbrücke Jena e. V., 2019), sind von großer Relevanz. Inwiefern das STT in der Lage ist, das selbstgenannte Ziel des Ausgleichs sozialer Benachteiligung zu erreichen (Kindersprachbrücke Jena e. V., 2019), verbleibt nach der dreijährigen Evaluation fraglich. Dass es sich bei den an der Evaluation beteiligten Schüler:innen um eine diverse und vulnerable Stichprobe handelt, ist aus den demographischen Hintergründen der Schüler:innen deutlich abzulesen (siehe Abschnitt 2.1.2). Zudem ist an den zahlreichen signifikanten Effekten für die zeitliche Entwicklung zentraler Variablen (bspw. bei der Motivation, beim Engagement) sowohl in der STT- aber auch in der Kontrollgruppe ablesbar, dass es sich bei der hier betrachteten Klassenstufen (5.-8. Klasse) um Schüler:innen in einer Entwicklungsphase handelt, in der sich ihr schulisches Erleben und Verhalten verändert. **Wir empfehlen daher, die Vulnerabilität der Schüler:innen im Fokus zu behalten. Mag das STT (bisher) keine positiven Effekte aufweisen, sollte eine langfristige Begleitung und Förderung insbesondere von sozial benachteiligten Schüler:innen der Schulklassen 5 bis 8 durch geeignete Maßnahmen beibehalten werden.**

6.4 Abschlussbewertung

Im Rahmen der ersten Förderperiode (Mitte et al., 2023) sowie am Rande der Evaluation wurde von zahlreichen Beteiligten zurückgemeldet, dass das STT erhebliche positive Einflüsse auf den Schulablauf, die schulinternen wie -externen Kommunikation und Organisationsabläufe hat. Teamteachende berichteten, dass durch ihr Mitwirken in der Schule bei Schulleitungen ein Bewusstsein und eine Öffnung hin zu schulförderlichen, unterstützenden Maßnahmen entstanden sei. Es wurde angedeutet, dass das STT tieferliegende, förderliche Veränderungsprozesse angestoßen habe, deren positiven Auswirkungen erst im weiteren Zeitverlauf zu messbaren Ergebnissen führen würde.

Wie oben ausführlich dargestellt, können bei mehrjährigem Vergleich von Klassen mit und ohne STT jedoch keine positiven Veränderungen bei den Schüler:innen festgestellt werden. Im Gegenteil wurde in vielen Zielbereichen (Emotionsregulation, Verhalten in der Klasse, Motivation, Engagement, Absentismus und Noten) festgestellt, dass Schüler:innen in Klassen mit STT über die Zeit schlechter abschnitten als jene in Klassen ohne STT.

Angenommen, dass die multiplen Krisen der letzten Jahre (bspw. Corona-Pandemie, Ukraine-Krieg) sowie Herausforderungen durch einen tiefgreifenden demographischen Wandel nicht abebben werden, ist es umso wichtiger, in Maßnahmen zu investieren, die nicht nur langfristig positiv schulische Strukturen beeinflussen, sondern auch zeitnah und konsistent dem Bildungserfolg der Schüler:innen zugutekommen. Vor dem Hintergrund der Evaluationsergebnisse sind Entscheidungsträger:innen daher angehalten zu prüfen, wie das STT

grundlegend überarbeitet werden kann oder ob alternative Maßnahmen notwendig sind, um negative Auswirkungen auf Schüler:innen zu vermeiden und ihre schulische Entwicklung wirksam zu fördern.

6.5 Das Wichtigste auf einen Blick

Datenerhebung

- 3 Messzeitpunkte mit STT- und Kontrollgruppen zwischen 02/23 und 12/25
- 16 Thüringer Schulen mit insgesamt 50 Schulklassen
- Daten von 882 Schüler:innen, 25 Teamteachenden und 44 Klassenleitungen

Weil sich die Verteilung der Klassen zwischen der STT- und Kontrollgruppe mehrfach über den Erhebungszeitraum hinweg veränderte, wurden separate Analysen nach zwei Ansätzen gerechnet:

- *intention to treat*: ursprünglich geplante Zuteilung
- *per protocol*: nur Teilnehmende, die die Evaluation gemäß den Vorgaben vollständig durchlaufen haben

Beide Analyse-Ansätze zeigen eine durchweg negative Entwicklung durch das STT:

- 11 (ITT, PP: 10) der 22 Ergebnismaße der Fragebogenerhebung entwickeln sich signifikant negativ über die Zeit
- für 10 Ergebnismaße finden sich Nulleffekte (ITT, PP: 11)
- Schüler:innen der STT-Gruppe haben signifikant mehr unentschuldigte Fehltage und signifikant schlechtere Noten in den Fächern Deutsch, Mathematik und Englisch als die Schüler:innen der Kontrollgruppe
- allgemein hohe Werte bei Zielvariablen der Lehrkräfte und Teamteachenden, u.a. Arbeitszufriedenheit und Identifikation mit der Schule, nur ein signifikanter Effekt: Stressbelastung der Lehrkräfte höher in der STT-Gruppe

Prüfung von Erklärungsansätzen

- Selektionseffekte oder Stichprobencharakteristika konnten ausgeschlossen werden
- am ehesten lassen sich die negativen Effekte durch eine Einschränkung des Autonomiebedürfnisses der Schüler:innen der STT-Klassen aufklären

6.6 Empfehlungen auf einen Blick

Stärkerer Fokus auf das Autonomieerleben der Schüler:innen

Wir empfehlen, dass die offenbar als vertrauensvoll erlebte Beziehung zu den Teamteachenden einen Rahmen bietet, um den erlebten Autonomieverlust zu kompensieren. So sollten Selbstkompetenzen gefördert werden, die gleichzeitig die Eigenständigkeit der Schüler:innen bspw. selbstverantwortliche Vorbereitung auf den Unterricht oder den Erwerb von Lernkompetenzen fördern.

Intensität und Dauer des STT zielgerechter gestalten

Wir empfehlen, das STT zumindest innerhalb der Schulen nicht nach wechselnden Kriterien einzelnen Klassen zuzuordnen, sondern bspw. für einen festen Zeitraum einer Jahrgangsstufe zuzuordnen oder einem konkreten Schulfach. Dies könnte insbesondere in jüngeren Klassenstufen oder in Übergangsphasen hilfreich sein.

Rollenklarheit fördern

Wir empfehlen, nicht nur den STT-Tandems, sondern auch dem gesamten Kollegium konkrete Handlungsrahmen vorzulegen, welche Rolle und Aufgabenfelder die Teamteachenden in ihrer Klasse und an der jeweiligen Schule übernehmen können/sollen – und welche genau nicht.

Wirkebenen des STT reflektieren

Wir empfehlen erstens, die Ziele des STT deutlich klarer und wissenschaftlich fundierter zu definieren sowie zu berücksichtigen, dass Ziele verschiedener Wirkungsebenen sich gegenseitig beeinträchtigen könnten. Zweitens sollte transparent sein, ob der Schwerpunkt auf der individuellen Förderung sozio-emotionaler Kompetenzen liegt oder ob eine breitere Veränderung schulischer Haltungen und Strukturen im Sinne einer Schulentwicklung angestrebt wird.

Fokus auf Vulnerabilität der Schüler:innen

Wir empfehlen, die Vulnerabilität der Schüler:innen im Fokus zu behalten. Mag das STT (bisher) keine positiven Effekte aufweisen, sollte eine langfristige Begleitung und Förderung insbesondere von sozial benachteiligten Schüler:innen der Schulklassen 5 bis 8 durch geeignete Maßnahmen beibehalten werden.

Literaturverzeichnis

- Abler, B. & Kessler, H. (2009). Emotion Regulation Questionnaire – Eine deutschsprachige Fassung des ERQ von Gross und John. *Diagnostica*, 55(3), 144–152.
<https://doi.org/10.1026/0012-1924.55.3.144>
- Akiba, M., Murata, A., Howard, C. C. & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, 77, 352–365. <https://doi.org/10.1016/j.tate.2018.10.012>
- Alivernini, F. & Lucidi, F. (2011). Relationship between social context, self-efficacy, motivation, academic achievement, and intention to drop out of high school: A longitudinal study. *The Journal of Educational Research*, 104(4), 241–252.
<https://doi.org/10.1080/00220671003728062>
- Álvarez, I., Fuertes Gutiérrez, M. & Gallardo Barbarroja, M. (2024). Team teaching in languages: A scoping review of approaches and practices in higher education. *Innovation in Language Learning and Teaching*, 18(1), 59–77. <https://doi.org/10.1080/17501229.2023.2229798>
- Anwar, K., Asari, S., Husniah, R. & Asmara, C. H. (2021). Students' perceptions of collaborative team teaching and student achievement motivation. *International Journal of Instruction*, 14(1), 325–344. <https://eric.ed.gov/?id=EJ1282190>
- Berry, B., Johnson, D. & Montgomery, D. (2005). The power of teacher leadership. *Educational Leadership*, 62(5), 56–60.
- Bolam, R., McMahan, A., Stoll, L., Thomas, S. & Wallace, M. (2005). *Creating and sustaining effective professional learning communities*. University of Bristol.
- Böttcher, W., Brockmann, L., Meierjohann, T. & Wiesweg, J. (2022). Was brauchen Schulen in herausfordernden Lagen? Eine explorative Studie im Auftrag des Netzwerk Bildung der Friedrich-Ebert-Stiftung. *FES diskurs*. <https://library.fes.de/pdf-files/a-p-b/19077.pdf>
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. *Soziale Welt Sonderband*, 2, 183–198.
- Chitiyo, J. (2017). Challenges to the use of co-teaching by teachers. *International Journal of Whole Schooling*, 13(3), 55–66.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.
- Deci, E. L., La Guardia, J. G., Moller, A. C., Scheiner, M. J. & Ryan, R. M. (2006). On the benefits of giving as well as receiving autonomy support: Mutuality in close friendships. *Personality & social psychology bulletin*, 32(3), 313–327.
<https://doi.org/10.1177/0146167205282148>

- Decuyper, A., Tack, H., Vanblaere, B., Simons, M. & Vanderlinde, R. (2023). Collaboration and shared responsibility in team teaching: A large-scale survey study. *Education Sciences*, 13(9), 896. <https://doi.org/10.3390/educsci13090896>
- Dishion, T. J. & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology*, 62, 189–214. <https://doi.org/10.1146/annurev.psych.093008.100412>
- Ditton, H. & Maaz, K. (2022). Sozioökonomischer Status, Bildungserfolg und Bildungsteilhabe. In H. Reinders, D. Bergs-Winkels, A. Prochnow & I. Post (Hrsg.), *Empirische Bildungsforschung* (S. 1083–1103). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27277-7_57
- Dodge, K. A., Dishion, T. J. & Lansford, J. E. (2006). Deviant peer influences in intervention and public policy for youth. *Social Policy Report*, 20(1), 1–20. <https://doi.org/10.1002/j.2379-3988.2006.tb00046.x>
- Duke, D. L., Tucker, P. D. & Salmonowicz, M. J. (2014). *Teachers' guide to school turnarounds* (2. Aufl.). Rowman & Littlefield Publishers.
- Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, 41(3-4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Ebokaiwe, N. T., Ajaja, O. P. & Kpangban, E. (2024). Effect of 5E guided inquiry, team teaching, and lecture method on physics students academic achievement in secondary schools in Delta State. *International Journal of Education, Training and Learning*, 8(1), 1–5. <https://ideas.repec.org/a/oap/ijetal/v8y2024i1p1-5id1346.html>
- Fend, H. & Prester, H.-G. (2014). *Fragebogenerhebung Entwicklung im Jugendalter (Skalenkollektion)*. <https://doi.org/10.7477/43:47:1>
- Forsyth, P. B., Adams, C. M. & Hoy, W. K. (2011). *Collective trust: Why schools can't improve without it*. Teachers College Press.
- Fredricks, J. A. & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of research on student engagement* (S. 763–782). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_37
- Froiland, J. M., Oros, E., Smith, L. & Hirschert, T. (2012). Intrinsic motivation to learn: The nexus between psychological health and academic success. *Contemporary School Psychology*, 16, 91–100. <https://doi.org/10.1007/BF03340978>

- Furrer, C. J., Skinner, E. A. & Pitzer, J. R. (2014). The influence of teacher and peer relationships on students' classroom engagement and everyday motivational resilience. *National Society for the Study of Education*, 113(1), 101–123.
<https://doi.org/10.1177/016146811411601319>
- Gentrup, S., Rjosk, C., Stanat, P. & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen. *Zeitschrift für Erziehungswissenschaft*, 21(4), 867–891. <https://doi.org/10.1007/s11618-018-0806-2>
- Glaesser, D., Holl, C., Malinka, J., McCullagh, L., Meissner, L., Harth, N. S., Machunsky, M. & Mitte, K. (2024). Examining the association between social context and disengagement: Individual and classroom factors in two samples of at-risk students. *Social Psychology of Education*(27), 115–150. <https://doi.org/10.1007/s11218-023-09829-4>
- Gnambs, T. & Hanfstingl, B. (2014). A differential item functioning analysis of the German Academic Self-Regulation Questionnaire for adolescents. *European Journal of Psychological Assessment*, 30(4), 251–260. <https://doi.org/10.1027/1015-5759/a000185>
- Gollwitzer, M. & Jäger, R. S. (2014). *Evaluation kompakt* (2., überarbeitete Auflage). Beltz.
- Grgic, M. & Bayer, M. (2015). Eltern und Geschwister als Bildungsressourcen? Der Beitrag von familialem Kapital für Bildungsaspirationen, Selbstkonzept und Schulerfolg von Kindern. *Zeitschrift für Familienforschung*, 27(2), 173–192.
- Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3), 281–291. <https://doi.org/10.1017/s0048577201393198>
- Gross, J. J. & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Gubbels, J., van der Put, C. E. & Assink, M. (2019). Risk factors for school absenteeism and dropout: A meta-analytic review. *Journal of youth and adolescence*, 48(9), 1637–1667. <https://doi.org/10.1007/s10964-019-01072-5>
- Hackman, J. R. & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *The Journal of applied psychology*, 60(2), 159–170. <https://doi.org/10.1037/h0076546>
- Haller, A. O. (1968). On the concept of aspiration. *Rural Sociology*, 33(4), 484–487.
- Härkki, T., Vartiainen, H., Seitamaa-Hakkarainen, P. & Hakkarainen, K. (2021). Co-teaching in non-linear projects: A contextualised model of co-teaching to support educational change. *Teaching and Teacher Education*, 97, 103188.
<https://doi.org/10.1016/j.tate.2020.103188>

- Hennemann, T., Hagen, T. & Hillenbrand, C. (2010). Dropout aus der Schule: Empirisch abgesicherte Risikofaktoren und wirksame pädagogische Maßnahmen. *Empirische Sonderpädagogik*(3), 26–47.
- Heppt, B., Olczyk, M. & Volodina, A. (2022). Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for academic achievement. *Social Psychology of Education*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s11218-022-09704-8>
- Herscovitch, L. & Meyer, J. P. (2002). Commitment to organizational change: Extension of a three-component model. *Journal of applied psychology*, 87(3), 474–487. <https://doi.org/10.1037//0021-9010.87.3.474>
- Holl, C., Gläßer, D., Schröder, A., Schulz, S., Mitte, K. & Harth, N. S. (accepted for publication). Schulabbrüche verstehen – Fachkräftepotenziale nutzen: Jugendliche mit Migrationshintergrund im Fokus. *Personalquarterly*.
- Hollins, E., McIntyre, L., DeBose, C., Hollins, K. & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International Journal of Qualitative Studies in Education*, 17(2), 247–264.
- Holtappels, H. G., Webs, T., Kamarianakis, E. & van Ackeren, I. (2017). Schulen in herausfordernden Problemlagen: Typologien, Forschungsstand und Schulentwicklungsstrategien. In V. Manitijs & P. Döbelstein (Hrsg.), *Beiträge zur Schulentwicklung. Schulentwicklungsbildung in herausfordernden Lagen* (1. Aufl., S. 17–35). Waxmann.
- Institut der Deutschen Wirtschaft Köln (Hrsg.). (2019). *INSM-Bildungsmonitor 2019: Ökonomische Bildung und Teilhabechancen*. Studie im Auftrag der Initiative Neue Soziale Marktwirtschaft (INSM).
- Institut für Qualitätsentwicklung Hessen. (2013). *Fragebogenerhebung Hessischer Referenzrahmen Schulqualität (Skalenkollektion)*. <https://doi.org/10.7477/44:48:1>
- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W. & Röder, B. (2009). *Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht: Skalen zur Erfassung von Lehrer- und Schülermerkmalen*. https://www.erziehungswissenschaften.hu-berlin.de/de/paedpsych/forschung/Skalenbuch_FoSS.pdf
- Johnson, R. R., Test, D. W. & Algozzine, B. (1995). Special education in general education classrooms: Cooperative teaching using supportive learning activities. *Special Services in the schools*, 10(1), 25–43.

- Jortveit, M. & Kovač, V. B. (2022). Co-teaching that works: Special and general educators' perspectives on collaboration. *Teaching Education*, 33(3), 286–300.
<https://doi.org/10.1080/10476210.2021.1895105>
- Jurkowski, S. & Müller, B. (2018). Co-teaching in inclusive classes: The development of multiprofessional cooperation in teaching dyads. *Teaching and Teacher Education*, 75, 224–231. <https://doi.org/10.1016/j.tate.2018.06.017>
- Jurkowski, S., Ulrich, M. & Müller, B. (2020). Co-teaching as a resource for inclusive classes: Teachers' perspectives on conditions for successful collaboration. *International Journal of Inclusive Education*, 1–18. <https://doi.org/10.1080/13603116.2020.1821449>
- Jurkowski, S., Ulrich, M. & Müller, B. (2023). Co-teaching as a resource for inclusive classes: teachers' perspectives on conditions for successful collaboration. *International Journal of Inclusive Education*, 27(1), 54–71. <https://doi.org/10.1080/13603116.2020.1821449>
- Kaempfe, S. & Krause, A. (2004). Gefährdungsbeurteilungen zur Analyse psychischer Belastungen am Arbeitsort Schule. In W. Bungard, B. Koop & C. Liebig (Hrsg.), *Psychologie und Wirtschaft leben: Aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (1. Aufl., S. 314–319). Hampp.
- Kaiser, L.-M., Besa, K.-S., Wilde, M. & Großmann, N. (2021). Eine mehrdimensionale Betrachtung des Druckerlebens von Schüler*innen der Sekundarstufe II aus der Perspektive der Selbstbestimmungstheorie. *Zeitschrift für Erziehungswissenschaft*, 24(6), 1401–1427. <https://doi.org/10.1007/s11618-021-01056-x>
- Kearney, C. A. (2008). School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical psychology review*, 28(3), 451–471.
<https://doi.org/10.1016/j.cpr.2007.07.012>
- Keppens, G., Spruyt, B. & Dockx, J. (2019). Measuring school absenteeism: Administrative attendance data collected by schools differ from self-reports in systematic ways. *Frontiers in psychology*, 10, 2623. <https://doi.org/10.3389/fpsyg.2019.02623>
- Kil, M., Leffelsend, S. & Metz-Göckel, H. (2000). Zum Einsatz einer revidierten und erweiterten Fassung des Job Diagnostic Survey im Dienstleistungs- und Verwaltungssektor. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 44(3), 115–128.
<https://doi.org/10.1026//0932-4089.44.3.115>
- Kindersprachbrücke Jena e. V. (2019). *Handbuch sozialpädagogisches Teamteaching* (4. Aufl.). Kindersprachbrücke Jena e. V.
- Kindersprachbrücke Jena e. V. (2024). *Handbuch Sozialpädagogisches Teamteaching*.
- Kipping, R. R., Howe, L. D., Jago, R., Campbell, R., Wells, S., Chittleborough, C. R., Mytton, J., Noble, S. M., Peters, T. J. & Lawlor, D. A. (2014). Effect of intervention aimed at

- increasing physical activity, reducing sedentary behaviour, and increasing fruit and vegetable consumption in children: active for Life Year 5 (AFLY5) school based cluster randomised controlled trial. *BMJ (Clinical research ed.)*, 348, g3256.
<https://doi.org/10.1136/bmj.g3256>
- Kleine, L. (2014). *Der Übergang in die Sekundarstufe I: Die Bedeutung sozialer Beziehungen für den Schulerfolg und die Formation elterlicher Bildungsentscheidungen*. Zugl. Bamberg, Univ., Diss., 2013. *Schriften aus der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität: Bd. 14*. Univ. of Bamberg Press.
- Kokko, M., Takala, M. & Pihlaja, P. (2021). Finnish teachers' views on co-teaching. *British Journal of Special Education*, 48(1), 112–132. <https://doi.org/10.1111/1467-8578.12348>
- Kristensen, T. S. & Borg, V. (2000). *AMI's spørgeskema om psykisk arbejdsmiljø*. National Institute of Occupational Health.
- Kristensen, T. S., Hannerz, H., Høgh, A. & Borg, V. (2005). *PsycTESTS Dataset*. American Psychological Association (APA). <https://doi.org/10.1037/t46049-000>
- Kuger, S., Jude, N., Klieme, E. & Kaplan, D. (2016). *PISA 2015 Field Trial – Chapter 4*.
<https://doi.org/10.7477/150:0:1>
- Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W. & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of personality and social psychology*, 95(1), 144–165. <https://doi.org/10.1037/0022-3514.95.1.144>
- Leahy, A. A., Eather, N., Smith, J. J., Hillman, C., Morgan, P. J., Nilsson, M., Lonsdale, C., Plotnikoff, R. C., Noetel, M., Holliday, E., Shigeta, T. T., Costigan, S. A., Walker, F. R., Young, S., Valkenborghs, S. R., Gyawali, P., Harris, N., Kennedy, S. G. & Lubans, D. R. (2019). School-based physical activity intervention for older adolescents: Rationale and study protocol for the Burn 2 Learn cluster randomised controlled trial. *BMJ open*, 9(5), e026029. <https://doi.org/10.1136/bmjopen-2018-026029>
- Lindner, K.-T. & Schwab, S. (2020). Differentiation and individualisation in inclusive education: A systematic review and narrative synthesis. *International Journal of Inclusive Education*, 1–21. <https://doi.org/10.1080/13603116.2020.1813450#:~:text=https%3A//doi.org>
- MacAulay, D. J. (1990). Classroom environment: A literature review. *Educational Psychology*, 10(3), 239–253. <https://doi.org/10.1080/0144341900100305>
- McCord, J. & Tremblay, R. E. (Hrsg.). (1992). *Preventing antisocial behavior: Interventions from birth through adolescence*. Guilford Press.
- Mitte, K., Harth, N. S., Machunsky, M., Casper, M., Glaesser, D., Holl, C., Malinka, J., McCullagh, L. & Schmidt, T. (2023). *Evaluierung des Sozialpädagogischen Tamteachings*

- von 2019-2022: Abschlussbericht. Ernst-Abbe-Hochschule Jena. https://www.sw.eah-jena.de/wp-daten/uploads/STT_Mitte_Harth_Abschlussbericht_Final_mit-Zitiervorschlag.pdf
- Morinaj, J. & Hascher, T. (2019). School alienation and student well-being: A cross-lagged longitudinal analysis. *European Journal of Psychology of Education*, 34(2), 273–294. <https://doi.org/10.1007/s10212-018-0381-1>
- Moser Opitz, E., Maag Merki, K., Pfaffhauser, R., Stöckli, M. & Garrote, A. (2021). Die Wirkung von unterschiedlichen Formen von co-teaching auf die von den Schülerinnen und Schülern wahrgenommene Unterrichtsqualität in inklusiven Klassen. *Unterrichtswissenschaft*, 49(3), 443–466. <https://doi.org/10.1007/s42010-021-00103-0>
- Murawski, W. W. & Swanson, H. L. (2001). A meta-analysis of co-teaching research: Where are the data? *Remedial and Special Education*, 22(5), 258–267. <https://doi.org/10.1177/074193250102200501>
- Nachtigall, C. (2008, 11. November). *Landesbericht: Thüringer Kompetenztests 2008*. Friedrich-Schiller-Universität Jena. <https://d-nb.info/1177633787/34>
- Nachtigall, C. (Hrsg.). (2018). *Landesbericht Thüringer Kompetenztests 2018*. Friedrich-Schiller-Universität Jena. https://www.kompetenztest.de/download/kt2018_landesbericht.pdf
- Nevin, A. I., Cramer, E., Voigt, J. & Salazar, L. (2008). Instructional modifications, adaptations, and accommodations of coteachers who loop. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 31(4), 283–297. <https://doi.org/10.1177/0888406408330648>
- Newmann, F. M. & Wehlage, G. G. (1995). *Successful school restructuring: A report to the public and educators*. American Federation of Teachers.
- Nouwen, W. & Clycq, N. (2020). Assessing the added value of the self-system model of motivational development in explaining school engagement among students at risk of early leaving from education and training. *European Journal of Psychology of Education*, 36, 243–261. <https://doi.org/10.1007/s10212-020-00476-3>
- Nübling, M., Vomstein, M., Haug, A. & Lincke, H.-J. (2015). *Erhebung psychosozialer Belastungen bei Lehrkräften im Rahmen von Gefährdungsbeurteilungen mit dem COPSOQ: Gesamtergebnis Regierungsbezirk Detmold (T5)*. FFAS: Freiburger Forschungsstelle Arbeits- und Sozialmedizin. <https://www.bezreg-arnsberg.nrw.de/themen/c/copsoq/COPSOQ-Gesamtergebnis-Regierungsbezirk-Arnsberg.pdf>

- Núñez, J. L. & León, J. (2015). Autonomy support in the classroom: A review from Self-Determination Theory. *European Psychologist*, 20(4), 275–283.
<https://doi.org/10.1027/1016-9040/a000234>
- Palinkas, L. A., Atkins, C. J., Miller, C. & Ferreira, D. (1996). Social skills training for drug prevention in high-risk female adolescents. *Preventive medicine*, 25(6), 692–701.
<https://doi.org/10.1006/pmed.1996.0108>
- Pekrun, R. (2016). Academic Emotions. In K. R. Wentzel & D. B. Miele (Hrsg.), *Handbook of Motivation at School* (2. Aufl., S. 120–144). Routledge.
- Rabin, C. (2020). Co-teaching: Collaborative and caring teacher preparation. *Journal of Teacher Education*, 71(1), 135–147. <https://doi.org/10.1177/0022487119872696>
- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolf, H.-G., Rost, J. & Schiefele, U. (2012). *Fragebogenerhebung PISA/II - Internationale Studie (2003) (Skalenkollektion)*. <https://doi.org/10.7477/15:287:1>
- Randall, R. & Marangell, J. (2021). Changing what we might have done on our own: Improving classroom culture and learning through teacher collaboration. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 94(1), 38–46.
<https://doi.org/10.1080/00098655.2020.1828240>
- Rauterberg, R. J. & Hauksdóttir, D. (2023). Gemeinsame Verantwortung und nach nachhaltige Unterstützung: Multiprofessionelle Teamarbeit in einer inklusiven Schule - eine Fallstudie. In N. Leonhardt, R. Kruschel, R. Jörgensdóttir Rauterberg & A. Hinz (Hrsg.), *Inklusive Bildung in Island: Grundlagen, Praktiken und Reflexionen* (1. Auflage). Juventa Verlag.
- Rea, P. J., McLaughlin, V. L. & Walther-Thomas, C. (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children*, 68(2), 203–223.
- Reeve, J. & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257–267.
<https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Rexroat-Frazier, N. & Chamberlin, S. (2019). Best practices in co-teaching mathematics with special needs students. *Journal of Research in Special Educational Needs*, 19(3), 173–183. <https://doi.org/10.1111/1471-3802.12439>
- Rosenthal, R. (2002). The Pygmalion Effect and its mediating mechanisms. In *Improving Academic Achievement* (S. 25–36). Elsevier. <https://doi.org/10.1016/B978-012064455-1/50005-1>

- Roth, J. & Mazziotta, A. (2015). Adaptation and validation of a German multidimensional and multicomponent measure of social identification. *Social Psychology*, 46(5), 277–290. <https://doi.org/10.1027/1864-9335/a000243>
- Ryan, R. M. & Deci, E. L. (2000). Self-Determination Theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Rytivaara, A., Pulkkinen, J. & Bruin, C. L. de (2019). Committing, engaging and negotiating: Teachers' stories about creating shared spaces for co-teaching. *Teaching and Teacher Education*, 83, 225–235. <https://doi.org/10.1016/j.tate.2019.04.013>
- Sälzer, C. C. (2009). *Schule und Absentismus: Zur Bedeutung schulischer Kontextfaktoren für die Erklärung und Vorhersage von Schulabsentismus* [Doktorarbeit]. Universität Freiburg, Freiburg.
- Schulze, A. & Preisendörfer, P. (2013). Bildungserfolg von Kindern in Abhängigkeit von der Stellung in der Geschwisterreihe. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 65(2), 339–356. <https://doi.org/10.1007/s11577-013-0205-x>
- Solis, M., Vaughn, S., Swanson, E. & McCulley, L. (2012). Collaborative models of instruction: The empirical foundations of inclusion and co-teaching. *Psychology in the Schools*, 49(5), 498–510. <https://doi.org/10.1002/pits.21606>
- Stefanidis, A., King-Sears, M. E., Strogilos, V., Berkeley, S., DeLury, M. & Voulagka, A. (2023). Academic achievement for students with and without disabilities in co-taught classrooms: A meta-analysis. *International Journal of Educational Research*, 120, 102208. <https://doi.org/10.1016/j.ijer.2023.102208>
- Stefanidis, A., King-Sears, M. E. & Brawand, A. (2019). Benefits for coteachers of students with disabilities: Do contextual factors matter? *Psychology in the Schools*, 56(4), 539–553. <https://doi.org/10.1002/pits.22207>
- Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal*, 104(2), 127–146. <https://doi.org/10.1086/499746>
- Strogilos, V., King-Sears, M. E., Tragoulia, E., Voulagka, A. & Stefanidis, A. (2023). A meta-synthesis of co-teaching students with and without disabilities. *Educational Research Review*, 38, 100504. <https://doi.org/10.1016/j.edurev.2022.100504>
- Takala, M. & Uusitalo-Malmivaara, L. (2012). A one-year study of the development of co-teaching in four Finnish schools. *European Journal of Special Needs Education*, 27(3), 373–390. <https://doi.org/10.1080/08856257.2012.691233>

- Tanner-Smith, E. E. & Wilson, S. J. (2013). A meta-analysis of the effects of dropout prevention programs on school absenteeism. *Prevention Science*, 14(5), 468–478.
<https://doi.org/10.1007/s11121-012-0330-1>
- Thompson, M. & Schademan, A. (2019). Gaining fluency: Five practices that mediate effective co-teaching between pre-service and mentor teachers. *Teaching and Teacher Education*, 86, 102903. <https://doi.org/10.1016/j.tate.2019.102903>
- Thüringer Ministerium für Bildung, Jugend und Sport. (2022). *Richtlinie des Freistaats Thüringen über die Gewährung von Mitteln aus dem Europäischen Sozialfonds Plus (ESF+) zur Steigerung der Chancengleichheit und Bildungsgerechtigkeit gemäß dem Programm ESF+ Thüringen 2021-2027 (ESF+ Schulförderrichtlinie)*. <https://www.esf-thueringen.de/downloads/richtlinien/esf-schulfoerderrichtlinie>
- Tsybulsky, D. & Muchnik-Rozanov, Y. (2019). The development of student-teachers' professional identity while team-teaching science classes using a project-based learning approach: A multi-level analysis. *Teaching and Teacher Education*, 79, 48–59.
<https://doi.org/10.1016/j.tate.2018.12.006>
- van den Broeck, A., Vansteenkiste, M., Witte, H. de, Soenens, B. & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the Work-related Basic Need Satisfaction scale. *Journal of Occupational and Organizational Psychology*, 83(4), 981–1002.
<https://doi.org/10.1348/096317909X481382>
- van Garderen, D., Stormont, M. & Goel, N. (2012). Collaboration between general and special educators and student outcomes: A need for more research. *Psychology in the Schools*, 49(5), 483–497. <https://doi.org/10.1002/pits.21610>
- Vangrieken, K., Dochy, F., Raes, E. & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, 15, 17–40.
<https://doi.org/10.1016/j.edurev.2015.04.002>
- Vembye, M. H., Weiss, F. & Hamilton Bhat, B. (2024). The effects of co-teaching and related collaborative models of instruction on student achievement: A systematic review and meta-analysis. *Review of Educational Research*, 94(3), 376–422.
<https://doi.org/10.3102/00346543231186588>
- Wagner, W., Helmke, A. & Rösner, E. (2011). *DESI - Abschlussbefragung (Skalenkollektion)*.
<https://doi.org/10.7477/6:219:1>
- Wang, M.-T. & Eccles, J. S. (2011). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of*

- Research on Adolescence*, 22(1), 31–39. <https://doi.org/10.1111/j.1532-7795.2011.00753.x>
- Wang, M.-T., Fredricks, J., Ye, F., Hofkens, T. & Linn, J. S. (2019). Conceptualization and assessment of adolescents' engagement and disengagement in school: A multidimensional school engagement scale. *European Journal of Psychological Assessment*, 35(4), 592–606. <https://doi.org/10.1027/1015-5759/a000431>
- Webs, T. (2016). Commitment-Typen bei Lehrkräften und ihre Zusammenhänge zu Lehrer*innen-Gesundheit und Schulkultur. In R. Strietholt, W. Bos & Holtappels, H. G., McElvany, N. (Hrsg.), *Jahrbuch der Schulentwicklung* (S. 277–305). Beltz/ Juventa.
- Weerd, D. de, Simons, M., Struyf, E. & Tack, H. (2024). Studying the effectiveness of team teaching: A systematic review on the conceptual and methodological credibility of experimental studies. *Review of Educational Research*, Artikel 00346543241262807. Vorab-Onlinepublikation. <https://doi.org/10.3102/00346543241262807>
- Zigmond, N. & Baker, J. M. (1990). Mainstream experiences for learning disabled students (Project MELD): Preliminary report. *Exceptional Children*, 57(2), 176–185.

ANHANG

Anhang 1: Publikationsverzeichnis

Vorträge

Glaesser, D., Jugert, P., & Harth, N. S. (2023, August 21). *Associations of intergroup understanding and inter-ethnic friendships: Preliminary findings from a sample of Thuringian secondary schools*. 4th Cultural Diversity, Migration, and Education Conference, Halle (Saale).

Glaesser, D., Harth, N. S., Machunsky, M., McCullagh, L. & Mitte, K. (2023, 12. September). *A social identity approach to school disengagement: Individual and classroom processes in two samples of at-risk students*. Fachgruppentagung Sozialpsychologie, Graz.

Gläßer, D., Holl, C., Malinka, J., Machunsky, M., Harth, N. S. & Mitte, K. (2024, 26. April). *Soziale Identität und Schulunlust: Individuelle und klasseninterne Prozesse in zwei Stichproben von Schulen in herausfordernden Lagen*. Tagung der Deutschen Gesellschaft für Soziale Arbeit (DGSA), Jena.

Glaesser, D., Jugert, P., & Harth, N. S. (2024, 17. September). *Associations of intergroup understanding and inter-ethnic friendships: Multilevel data from secondary schools with high rates of first generation immigrant students*. 53. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Wien.

Holl, C., Gläßer, D., Malinka, J., Mitte, K. & Harth, N. S. (2024, 26. April). *Prekäre Bildungstransformationen in Thüringen: Ein Überblick makrodimensionaler Prozesse*. Tagung der Deutschen Gesellschaft für Soziale Arbeit (DGSA), Jena.

Schröder, A., Pötschlag, D., Sauer, V., Schlötzer, K., Mitte, K. & Harth, N. S. (2024, 26. April). *Zusammenarbeit im multiprofessionellen Team: Rollenverständnis und pädagogische Haltung im Sozialpädagogischen Teamteaching*. Tagung der Deutschen Gesellschaft für Soziale Arbeit (DGSA), Jena.

Schroeder, A., Glaesser, D., Holl, C., McCullagh, L., Mitte, K. & Harth, N. S. (2024, 17. September). *Students trust in socio-pedagogical co-teachers stabilizes longitudinal decreases of engagement: Findings from highly diverse classrooms*. 53. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Wien.

Schröder, A., Gläßer, D., Holl, C., McCullagh, L., Mitte, K. & Harth, N. S. (2025, 29. Januar). *Vertrauen von Schüler:innen in Teamteacher:innen als Puffer für die Abnahme schulischen Engagements: Längsschnittergebnisse aus Schulen in herausfordernden Lagen*. 12. Kongress der Gesellschaft für empirische Bildungsforschung (GEBF), Mannheim.

Schulz, S., Gläßer, D., Holl, C., Schröder, A., Mitte, K. & Harth, N. S. (2025, 30. September). *Wirksamkeit von Sozialpädagogischem Teamteaching an Thüringer Schulen in herausfordernden Lagen*. Fachgruppentagung Pädagogische Psychologie, Jena. [angenommen, wegen Krankheit ohne Teilnahme]

Publikationen

Glaesser, D., Holl, C., Malinka, J., McCullagh, L., Meissner, L., Harth, N. S., Machunsky, M. & Mitte, K. (2024). Examining the association between social context and disengagement: Individual and classroom factors in two samples of at-risk students. *Social Psychology of Education*(27), 115–150. <https://doi.org/10.1007/s11218-023-09829-4>

Holl, C., Gläßer, D., Schröder, A., Schulz, S., Mitte, K. & Harth, N. S. (Accepted for publication). Schulabbrüche verstehen – Fachkräftepotenziale nutzen: Jugendliche mit Migrationshintergrund im Fokus. *PERSONALquarterly*.

Anhang 2: Deskriptive Werte relevanter Variablen in der Schüler:innen-Stichprobe

Tabelle 20 Deskriptive Werte relevanter Variablen in der Schüler:innen Stichprobe über die Zeit

Variable	T1			T2			T3		
	<i>M (SD)</i>	Spann- weite	ω	<i>M (SD)</i>	Spann- weite	ω	<i>M (SD)</i>	Spann- weite	ω
Skalen zum STT und zu Lehrpersonen									
Vertrauen zu TT	4.22 (0.78)	1-5	.92	4.11 (0.81)	1.1-5	.91	4.08 (0.81)	1.1-5	.93
Akademische Unterstützung durch TT	3.34 (0.84)	1-5	.85	3.35 (0.80)	1-5	.85	3.29 (0.83)	1.2-5	.87
Wahrgenommene Fairness TT	4.56 (0.83)	1-5	.80	4.55 (0.78)	1-5	.79	4.53 (0.79)	1-5	.81
Wahrgenommene Fairness LK	4.14 (1.04)	1-5	.77	4.06 (1.04)	1-5	.78	4.04 (1.03)	1-5	.79
Akademische Unterstützung durch TT	3.34 (0.85)	1-5	.85	3.35 (0.80)	1-5	.85	3.29 (0.83)	1.2-5	.87
Wahrgenommene Autonomieunterstützung durch LK	3.45 (0.99)	1-5	.89	3.33 (1.08)	1-5	.89	3.93 (0.96)	1-5	.92
Bedürfniserfüllung, Wohlbefinden und Emotionsregulation									
Bedürfnis nach Verbundenheit	3.36 (1.03)	1-5	.77	3.22 (1.03)	1-5	.80	3.32 (0.99)	1-5	.79
Bedürfnis nach Kompetenz	3.46 (0.69)	1-5	.71	3.39 (0.69)	1-5	.78	3.43 (0.71)	1.2-5	.80
Bedürfnis nach Autonomie	3.16 (0.71)	1-5	.50	3.04 (0.66)	1-4.8	.56	2.98 (0.65)	1-4.8	.59
Wohlbefinden im Schulalltag	5.97 (2.67)	0-10		5.76 (2.56)	0-10		5.70 (2.48)	0-10	
Emotionsregulation: Neubewertung	3.09 (0.85)	1-5	.83	3.02 (0.83)	1-5	.86	3.01 (0.79)	1-5	.83
Emotionsregulation: Unterdrückung	3.20 (0.87)	1-5	.69	3.12 (0.84)	1-5	.73	3.22 (0.84)	1-5	.75
Schulerfolg									
Kognitives Engagement	3.71 (0.78)	1-5	.80	3.64 (0.80)	1-5	.84	3.59 (0.79)	1-5	.84
Verhaltensbezogenes Disengagement	2.10 (0.67)	1-5	.72	2.05 (0.69)	1-5	.81	2.08 (0.70)	1-5	.81
Agentisches Engagement	2.78 (0.92)	1-5	.77	2.75 (0.92)	1-5	.83	2.75 (0.92)	1-5	.84
Intrinsische Motivation	3.17 (0.95)	1-5	.84	2.99 (0.90)	1-5	.86	2.84 (0.85)	1-5	.86
Extrinsische Motivation	3.22 (0.88)	1-5	.63	3.25 (0.81)	1-5	.58	3.25 (0.78)	1-5	.58
Identifizierte Motivation	3.91 (0.86)	1-5	.67	3.74 (0.86)	1-5	.70	3.66 (0.85)	1-5	.70
Soziales Klassengefüge									
Klassenklima	3.29 (0.92)	1-5	.82	3.09 (0.89)	1-5	.82	3.15 (0.89)	1-5	.84
Verhalten im Klassenverbund	3.31 (0.89)	1-5	.88	3.25 (0.89)	1-5	.90	3.14 (0.86)	1-5	.90

Anmerkungen. Die Skalen, die sich auf Teamteacher:innen (TT) bezogen, wurden nur in der STT-Gruppe erhoben.