



# Developmental relations between mathematics self-concept, interest, and achievement: A comparison of solo- and co-taught classes

Hans Lehikoinen<sup>1</sup> · Pertti Väisänen<sup>1</sup> · Sari Havu-Nuutinen<sup>1</sup> ·  
Kristiina Lappalainen<sup>2</sup> · Markku Niemivirta<sup>1</sup>

Received: 22 March 2023 / Accepted: 19 July 2024  
© The Author(s) 2024

## Abstract

This study investigated the role of co-teaching in the development of students' mathematics motivation and achievement. More specifically, we examined how sixth-grade students' ( $N=146$ ) mathematics self-concept and individual interest changed over one school year, how these changes were related to each other and to mathematics achievement, and, most importantly, whether they differed between co-teaching and solo-teaching conditions. The co-teaching condition included 70 students in three classes with mathematics taught by pairs of teachers, while the solo-teaching condition included 76 students in four classes with mathematics taught by individual class teachers. The design included three repeated measures of mathematics self-concept and interest as well as pre- and post-measures of mathematics test performance and teacher-rated mathematics grades. A series of latent growth curve analyses showed both self-concept and individual interest to decline over time, and these changes to be strongly correlated: as self-concept decreased, so did interest, and vice versa. The changes in self-concept and interest were independent of prior achievement and did not predict later achievement either. Students in the co-taught group received better grades at the end of the year, but no differences in the development of self-concept and individual interest were found between the teaching conditions. That is, co-teaching contributed to improvement in mathematics achievement, but this was not mediated by changes in mathematics motivation.

**Keywords** Co-teaching · Mathematics motivation · Mathematics self-concept ·  
Mathematics interest · Mathematics achievement

---

✉ Hans Lehikoinen  
hansl@uef.fi

<sup>1</sup> School of Applied Educational Science and Teacher Education, University of Eastern Finland, P.O. Box 111, Joensuu 80101, Finland

<sup>2</sup> School of Educational Sciences and Psychology, University of Eastern Finland, Joensuu, Finland

## Introduction

Research shows numeracy skills to be influential in many areas of life, including school attainment and dropout, employment, and even psychological well-being (Hakkarainen et al., 2016; Parsons & Bynner, 2005). Given that student motivation in mathematics seems to decline over the school years (Jacobs et al., 2002; Pinxten et al., 2014), education faces an important challenge: how to foster mathematics learning and motivation already during the early years of education, when the foundations of mathematics skills and attitudes are gradually formed. This global challenge is relevant also in Finland, where mathematics achievement, despite its high level by international comparison, has been steadily declining in recent years, while students report simultaneously relatively low mathematics-related motivation and attitudes (OECD, 2013; Mullis et al., 2016).

Co-teaching has been considered as a means of increasing teacher responsiveness and instructional quality, and providing better support for students' different learning needs (e.g., Friend & Cook, 2013; Villa et al., 2008). The promise of co-teaching lies within the increased student-teacher interaction and effective teaching practices, such as individualisation and systematic feedback (Friend & Cook, 2013; Sweigart & Landrum, 2015; Villa et al., 2008). The motivational implications here are of particular importance, as research shows both feedback and individualised support to reinforce students' positive self-concept (O'Mara et al., 2006; Lüdtke et al., 2005) and higher interest (Kiemer et al., 2015). In this study, we investigated the potential of co-teaching in the context of primary school mathematics. More specifically, using a unique design, we compared the development of sixth-grade students' mathematics motivation (i.e., self-concept and individual interest) and achievement between co- and solo-taught classes.

## Mathematics self-concept and individual interest

Research on mathematics achievement shows students' evaluations of both themselves in relation to mathematics, and of mathematics as a school subject, to play an important role in mathematics learning (Middleton & Spanias, 1999). Particularly self-concept and interest in mathematics seem to be of specific relevance (e.g., Marsh et al., 2005). Mathematics self-concept refers to an individual's perception and evaluation of their mathematics abilities (Bong & Skaalvik, 2003). These perceptions are formed through experiences of interacting with the learning environment, and are influenced by environmental reinforcements, past achievements, and significant others (Shavelson et al., 1976). How students judge their achievements is considered a major determinant of self-concept, and this judgement grounds on internal and external comparison (Marsh & Craven, 2000; Skaalvik & Skaalvik, 2002). Internal comparison refers to an individual's comparison of their current abilities or achievement to past achievement within the same subject or between different subjects (Marsh & Craven, 1997; Skaalvik & Skaalvik, 2002), while external comparison reflects the comparison of one's own achievement to others' achievement, partly through feedback from peers, parents, and teachers (Gniewosz et al., 2012; Marsh & Craven, 1997; Skaalvik & Skaalvik, 2002).

Interest is another important factor facilitating learning (Harackiewicz et al., 2016). Learning being different when interest is either present or absent was noted already by Dewey (1913) – interest can “catch and hold” our attention to approach learning in a

“whole-hearted way”. Interest is thereby characterised by increased attention, concentration, and affect that are specific to a person, object, activity, or subject, and can be further divided into situational and individual forms (Hidi, 2006; Kaplan & Patrick, 2016; Schiefele, 2009). Situational interest refers to a short-lived temporary state, where environmental cues and factors can catch and maintain the attention of the individual, thus facilitating motivation to act in a certain way (Schiefele, 2009). Individual interest, instead, reflects a relatively long-lasting affective evaluation-orientation towards a specific subject or object (Hidi & Ainley, 2002; Schiefele, 2009). Sustained and maintained situational interest may gradually lead to individual interest over a time, although it may require support from the environment (Hidi & Renninger, 2006). Existing individual interest may, in turn, facilitate the triggering and maintenance of situational interest (Nuutila et al., 2020; Schiefele, 2009; Tapola et al., 2013); a student with high individual interest in mathematics is more likely to enjoy and engage in mathematics tasks than a student with low individual interest.

Research shows mathematics self-concept to promote a range of positive educational outcomes, such as academic choices and aspirations (e.g., Marsh et al., 2005), and contribute to achievement in a reciprocal manner (e.g., Marsh et al., 2005; Marsh & Craven, 2006; Marsh & Martin, 2011). That is, prior self-concept may boost later achievement, which, in turn, is likely to enhance subsequent self-concept. However, this reciprocity may not be as evident with younger students (Ganley & Lubienski, 2016; Viljaranta et al., 2014), as their self-concept tends to be inflated due to unrealistic appraisals (Wigfield & Eccles, 2000). Interest, instead, has been shown to facilitate attention as well as cognitive and affective processing (Ainley et al., 2002), and to predict effort (Arens & Hasselhorn, 2015) and course selection (Köller et al., 2001). High interest may also act as a buffer against the negative effects of non-optimal learning conditions (Hidi & Renninger, 2006; Katz et al., 2006; Tsai et al., 2008). Consequently, while self-concept seems to be more directly linked with achievement outcomes (e.g., Marsh et al., 2005), individual interest may serve as the mediating fuel to act towards learning as described by Dewey (1913).

There is also some evidence for self-concept and interest being interrelated both concurrently (Marsh et al., 2005) and longitudinally (Petersen & Hyde, 2017). That is, when students feel competent in mathematics, they are also more likely to experience interest and enjoyment in mathematics-related activities. However, the linkage between the two seems complex; previous findings provide support for both reciprocal relationships (Marsh et al., 2005) and the causal predominance of self-concept (Jacobs et al., 2002; Viljaranta et al., 2014; Wigfield & Eccles, 2000).

Research further indicates that students' mathematics self-concept (Jacobs et al., 2002; Petersen & Hyde, 2017; Wigfield & Eccles, 2000) and interest (Lazarides et al., 2019; Petersen & Hyde, 2017; Renninger & Hidi, 2011) decrease as students get older, although the level of interest may plateau across the later school years (Frenzel et al., 2010). It has been suggested that these developments, especially after mid-primary grades, are linked with more realistic appraisals of ability (Wigfield & Eccles, 2000), increased social comparison processes, and changes in instructional practices and curriculum (Stipek & Iver, 1989). Most prior research has focused on long-term changes over several years (e.g., Denissen et al., 2007), while less is known about the short-term (e.g., within a school year) developmental dynamics between mathematics self-concept and interest, and how these dynamics are connected with the pedagogical context and changes in achievement. Addressing these questions is one of the main aims and contributions of this study.

## Supporting self-concept and interest in mathematics learning

Despite the consistent observation of developmental decline in mathematics motivation, research also shows that students' mathematics self-concept (Watson et al., 2019; for a review see, O'Mara et al., 2006) and interest (Høgheim & Reber, 2015; Rotgans & Schmidt, 2017) can be supported through pedagogical practices. Teachers play an important role in the development of students' self-concept, as their actions and communication in the classroom provide the students with information for external and internal comparisons (e.g., Skaalvik & Skaalvik, 2002). One example is the teacher's use of social frame of reference standard, meaning that the teacher evaluates student achievement by a comparison between students, which may prompt students to view learning as competition (Lüdtke et al., 2005). Opposed to this is the teacher's use of individual frame of reference standard, whereby students' achievement is compared to their own past achievement, which, in turn, may facilitate personal effort and diminish social comparison (Lüdtke et al., 2005). Further, conveying high expectancies on students while downplaying social comparison may also help to maintain mathematics self-concept (Watson et al., 2019). Teacher feedback focusing on effort, improvement over time, and individual performance is likely to have a positive impact not only on students' mathematics self-concept (Lüdtke et al., 2005; O'Mara et al., 2006) but also on achievement emotions (Pekrun, 2016) and interest (Kaplan & Partick, 2016; Kiemer et al., 2015). In contrast, uninformative and controlling feedback (Deci et al., 1999; Tsai et al., 2008), and feedback endorsing social comparison or competition, may not only impede learning but also undermine intrinsic motivation (including interest) and threaten positive self-evaluations (Brophy, 2011; Marsh & Craven, 1997; Pekrun, 2016; Ryan & Deci, 2017).

Student interest can be further supported through interest-provoking didactic means such as introducing specific learning problems (Rotgans & Schmidt, 2017). Personalisation (Bernacki & Walkington, 2018; Høgheim & Reber, 2015) and differentiation seem to be the key issues here. For example, Durik et al. (2015) found that providing students with utility value information boosted their mathematics interest, but this was mostly the case for those with high success expectancies. In contrast, students with lower expectations benefited more from encouraging feedback that stressed effort and the belief that they possessed the required potential to perform well. Although such targeted pedagogical practices might seem rather self-evident, they do require effort and time. Personalising learning tasks for each student in order to spark their interest in a classroom full of students is no simple task (e.g., Hidi, 1990).

## Co-teaching as a means to improve instructional practices

One potential way to organise teaching to facilitate previously mentioned instructional practices and meet the students' needs better is co-teaching. This is a collaborative effort of two or more educators to combine their expertise to teach a heterogeneous class, usually within the same physical space, where they share all aspects of teaching, planning, and evaluation (Friend & Cook, 2013; Villa et al., 2008). The main goal is to provide an educational environment for all students to learn and succeed through an emphasis on such effective instructional practices that cannot be produced by a single teacher alone (Friend, 2008).

Co-teaching has been referred to, often erroneously, by many names, including collaborative teaching or team teaching. Co-teaching differs from team teaching by pairing profes-

sionals of different expertise within one class as opposed to two or more teachers pooling their classes into one, thus resulting in smaller teacher-student ratio (t-s ratio) (Conderman & Hedin, 2015; Friend et al., 2010; Sweigart & Landrum, 2015; Villa et al., 2008). It has been suggested that when co-teachers have more time to interact with their students, this may lead to not only increased scaffolding and individual support, but also more frequent and thorough feedback (Friend & Cook, 2013; Sweigart & Landrum, 2015).

Co-teaching thus involves much more than just adding adults in the classrooms (Sweigart & Landrum, 2015; Villa et al., 2008) or reducing class sizes. Pairing up with another teacher can offer several advantages to both teaching and planning. For instance, teachers may be able to identify and support different learners better, as some students benefit more from feedback, and some from teacher modelling and scaffolding (Kaplan & Patrick, 2016). When a teacher has the time to reflect on their instructional processes with another professional, it may lead to a reflective stance on their own professionalism that can enrich and further improve educational practices (Brophy, 1983; Rytivaara & Kershner, 2012). This co-teacher discourse and planning can break some ritualised routines that might hold back student learning (Nuthall, 2005), in contrast to a solo teacher who may be more restricted to internal dialogue, where self-justification might be more likely than critical reflection (Bright, 1996). It might not be about having the necessary skill-set, but to see the need for change. The dialogue as well as the increased student support may also shift the assessment of learning from the traditional end-of-period assessment to a daily routine that truly facilitates the teachers' individual frame of reference standard (e.g., Lüdtke et al., 2005).

Although the idea of co-teaching is appealing, research-based evidence on how it contributes to student outcomes is limited, and has often focused on students with special education needs (see, Cook et al., 2017; Strogilos et al., 2023). Murawski and Swanson's (2001) meta-analysis found co-teaching to be moderately effective in terms of learning outcomes. However, firm conclusions could not be drawn, since only six of the 89 reviewed studies were eligible for the actual analysis: some studies lacked comparison groups, and in some, the treatment fidelity was not adequately documented (Murawski & Swanson, 2001). One recent study showed co-teaching to impact students' test scores in mathematics positively (Jones & Winters, 2022). In the context of English classes, students reported having better access to individualised assistance, and co-teaching to contribute positively to their improvement (Wilson & Michaels, 2006). Lochner et al. (2019) compared solo- and co-taught classes using observational data, and found co-taught students to be more cognitively engaged in their learning, a finding similar to that of class-size reduction studies (Blatchford et al., 2011; Finn et al., 2003). This finding seems particularly relevant for the present study: could such an increase in engagement also boost student motivation?

## Present study

Although the research on mathematics self-concept, individual interest, and their development over time is relatively rich, we know less about how changes in them are connected with each other, and, particularly, how they correlate with changes in mathematics learning and achievement. Also, given how instructional practices found to support student learning and motivation (e.g., individualised tasks, feedback, support, and time to interact with different students) have been associated with co-teaching (Villa et al., 2008), surprisingly

little research has been conducted on its relative effectiveness. This would seem particularly relevant in the context of challenging, task intensive, and error-prone school subjects such as mathematics.

In this study, we investigated, (RQ1) how mathematics self-concept and individual interest change over one school year, and how these changes are related to each other, (RQ2) whether the levels and changes in self-concept and individual interest are predicted by the teaching condition (co-teaching versus solo-teaching), after controlling for teacher-rated achievement and test performance, and most importantly, (RQ3) how the teaching condition as well as changes in self-concept and interest further predict later achievement and test performance (i.e., teacher-rated grades and test performance at the end of sixth grade). As previous research has found some relatively consistent gender differences in mathematics motivation (Ganley & Lubienski, 2016; Jacobs et al., 2002), we also accounted for this by including gender as a covariate. To our knowledge, no previous studies have examined these developmental dynamics in the context of co-teaching.

Regarding the first research question, we expected both self-concept and interest to decline over time (Denissen et al., 2007; Frenzel et al., 2010; Jacobs et al., 2002; Pinxten et al., 2014), and these changes to be correlated (Petersen & Hyde, 2017).

As to the second research question, we anticipated co-teaching (as compared to solo teaching) to have a positive effect on the change in students' self-concept and interest due to more personalised teaching and pedagogical practices (e.g., Villa et al., 2008), while we assumed an effect of mathematics achievement and test performance on the onset of both self-concept and interest (Marsh et al., 2005; Ganley & Lubienski, 2016; Petersen & Hyde, 2017) as well as their changes (Denissen et al., 2007). That is, high achieving students were expected to have not only more positive self-concept and interest at the beginning of the sixth grade, but also show more positive changes in them.

Regarding the third research question, we expected both the initial level (Marsh et al., 2005) and relative improvement (or less steep decline) in self-concept and interest to contribute to better mathematics achievement and test performance by the end of the year (Denissen et al., 2007; Petersen & Hyde, 2017), after controlling for the effects of prior achievement. Most importantly, changes in self-concept and interest were in turn presumed to mediate the anticipated positive effect of co-teaching on mathematics achievement (e.g., Marsh et al., 2005; Marsh & Martin, 2011).

## Method

### Design

This study could be characterised as a natural experiment with some features of a quasi-experiment (Remler & Van Ryzin, 2011; Shadish et al., 2002). The initiative and design for the study came from the researchers, but the implementation was organised together with the participating schools and teachers. Following this, teachers were recruited to both conditions, and the practices to be implemented by the "experimental group" (i.e., co-teaching) were facilitated through workshops (for more details, see below). However, the co-teachers were completely free to plan and carry out their classes without any moderation or intervention by the researchers. In this sense, the events and activities that took place in both

conditions over the school year were naturally occurring and ecologically valid, despite the experimental setup (see, Bronfenbrenner, 1977).

The empirical arrangement followed a repeated-measures design with a comparison of two teaching conditions: (1) the co-teaching group, where mathematics was taught by pairs of class teachers and special education teachers, and (2) the solo teaching group, where mathematics was taught by individual class teachers who received some assistance from a special education teacher, which is common policy in Finland<sup>1</sup>. The students had three lessons of mathematics (45 min each) per week in both teaching conditions. The study design is illustrated in Fig. 1.

## Participants and procedure

Four primary schools and ten teachers within a city in Eastern Finland participated in the study. The student participants (Table 1) were 146 sixth grade students, aged 12–13 years. The co-teaching condition consisted of three classes (class size ranging from 23 to 27) for a total of 70 students; 47 girls and 23 boys. The solo-teaching condition consisted of four classes (class size ranging from 21 to 23) totalling 76 students; 40 girls and 36 boys. Gender distribution was relatively equal in both conditions,  $\chi^2(1)=3.19, p=.09$ .

The data were collected by questionnaires at three timepoints, as shown in Fig. 1: at the beginning of the school year in August (T1), at mid-term in January (T2), and in April (T3). Mathematics performance was tested at the beginning and at the end of the school year. The minor loss of participants at different timepoints was due to normal school absence. All data were collected by the first author in the classrooms, with the exception of T3 measure, which was conducted during an online lesson in April due to Covid-19 restrictions.

The recruitment began by contacting the school principals, who forwarded the request to participate to their teachers. The participating teachers had to have a Master's degree in education and sufficient teaching experience (i.e., minimum of five years), and they were required to join the project voluntarily (i.e., to not be assigned by administrators), to ensure sufficient comparability of participant background, as recommended by previous research (e.g., Friend & Cook, 2013; Saloviita & Takala, 2010; Scruggs et al., 2007). We also allowed teachers to form their co-teaching pairs within their respective schools by themselves, to

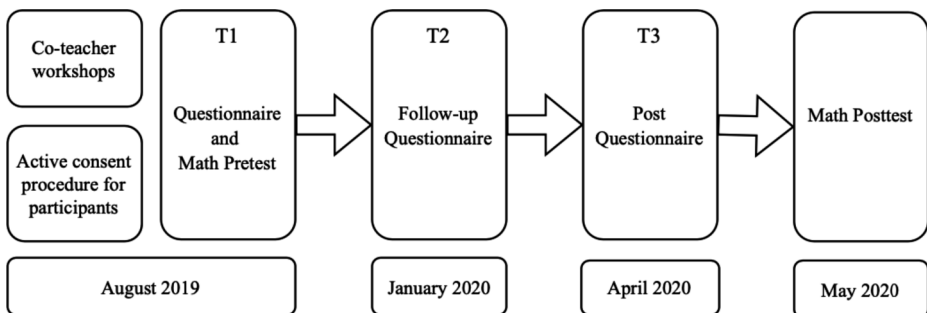


Fig. 1 Study design

<sup>1</sup> More than a fifth of students in Finland receive part-time special education (OSF, 2021). This flexible support system is for students who need additional help and does not require any diagnosis (see also, Savolainen et al., 2018).

**Table 1** Descriptive statistics and bivariate correlations for all variables

	Solo teaching						Co-teaching						All students												
	M		SD		N		M		SD		N		M		SD		N								
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	
1. Self-concept t1	3.9	0.74	76	3.8	0.76	70	3.85	0.75	-0.91	1.59	0.91	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. Self-concept t2	3.77	0.82	72	3.69	0.77	64	3.73	0.79	-0.87	1.45	0.90	0.80**	—	—	—	—	—	—	—	—	—	—	—	—	—
3. Self-concept t3	3.81	0.83	72	3.76	0.81	70	3.79	0.82	-1.22	1.95	0.92	0.74**	0.85**	—	—	—	—	—	—	—	—	—	—	—	—
4. Interest t1	3.47	1.06	76	3.39	1.02	70	3.43	1.04	-0.34	-0.68	0.96	0.54**	0.46**	0.43**	—	—	—	—	—	—	—	—	—	—	—
5. Interest t2	3.28	1.08	72	3.20	1.00	64	3.25	1.04	-0.10	-0.61	0.95	0.52**	0.54**	0.53**	0.78**	—	—	—	—	—	—	—	—	—	—
6. Interest t3	3.10	1.07	72	3.13	1.11	70	3.11	1.09	-0.29	-0.64	0.96	0.46**	0.50**	0.53**	0.68**	0.84**	—	—	—	—	—	—	—	—	—
7. Math test Pre	8.34	2.95	76	7.73	2.77	70	8.05	2.87	-0.02	-0.45	0.85	0.47**	0.51**	0.49**	0.29**	0.27**	0.26**	—	—	—	—	—	—	—	—
8. Math test Post	9.93	2.89	69	9.35	3.39	69	9.64	3.15	-0.33	-0.02	0.87	0.52**	0.53**	0.51**	0.28**	0.26**	0.27**	0.79**	—	—	—	—	—	—	—
9. Math grade Pre	8.54	0.94	76	8.38	1.01	65	8.47	0.98	-0.40	-0.04	0.59**	0.67**	0.61**	0.27**	0.33**	0.32**	0.67**	0.66**	—	—	—	—	—	—	—
10. Math grade Post	8.53	1.05	76	8.66	0.90	70	8.59	0.98	-0.34	-0.33	0.57**	0.66**	0.62**	0.29**	0.30**	0.32**	0.61**	0.64**	0.82**	—	—	—	—	—	—
11. Gender											0.01	-0.02	-0.01	-0.05	-0.07	-0.11	-0.11	-0.13	-0.21*	-0.15†	—	—	—	—	—
12. Group											-0.06	-0.05	-0.03	-0.04	-0.04	0.01	-0.11	-0.09	-0.08	0.07	-0.15†	—	—	—	—

Note. †  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

Gender coding: 0 = girls, 1 = boys. Group coding: 0 = solo teaching, 1 = co-teaching



facilitate more effective and equal partnerships with matching interests and pedagogical views (Pratt, 2014). All participants signed a written consent that followed the ethical guidelines of The Finnish National Board on Research Integrity, the university guidelines, and European Union GDPR requirements. According to the Finnish National Board on Research Integrity (2019) guidelines, no ethical review was necessary.

As many co-teaching studies suggest lack of knowledge to be one of the main obstacles to successful implementation of co-teaching (Friend et al., 2010; Saloviita & Takala, 2010; Scruggs et al., 2007), the project began with a workshop for each co-teacher pair. The workshop was led by one of the researchers, and included material on the basics and characteristics of co-teaching (Friend & Cook, 2013; Villa et al., 2008), teacher roles, as well as what is required for successful co-teaching (Murawski & Lochner, 2011; Pratt, 2014; Scruggs et al., 2007). The key aspects of co-teaching were emphasised: planning, teaching, and evaluation are to be shared, and these practices should include thorough reflection (Fluijt et al., 2016; Murawski & Lochner, 2011). Note, that the workshops did not include any material on self-concept or individual interest, or how to support them through teaching. The solo teachers did not receive any training from the researchers.

## Measures

### Mathematics self-concept and interest

The Self Description Questionnaire I (Marsh, 1990) was used to assess mathematics self-concept and individual interest. It is considered to be a reliable and valid instrument across different cultures (Arens & Hasselhorn, 2015), and has been validated in the Finnish context and in this age group (e.g., Savolainen et al., 2018). We followed recommendations to divide the scale into two components, competence self-perceptions and affect perceptions, respectively (Arens et al., 2011). Given the wording of the items within both components, we considered them as accurate representations of mathematics self-concept (i.e., the students' subjective evaluation of their competence in mathematics: "I am good at mathematics", "I find mathematics easy", "I learn mathematics quickly", and "I get good marks in mathematics"), and mathematics interest (i.e., students' affective and cognitive appraisals: "I look forward to mathematics", "I am interested in mathematics", "I like mathematics", and "I enjoy mathematics"). The students rated each item with a Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Both measures showed excellent reliability for all timepoints (T1-T3): the respective McDonald's omegas ( $\omega$ ) for self-concept were 0.91, 0.90, and 0.92, and for interest 0.96, 0.95, and 0.96.

### Mathematics performance and achievement

Students' mathematics performance was measured twice (at the beginning and at the end of sixth grade) with the standardised RMAT-test (Räsänen, 2004), which is a time-constrained test (i.e., 10 min) for basic numeracy skills in mathematics. The test comprises a set of basic arithmetic tasks (56 items): addition, subtraction, multiplication, division, fraction, units of measurement, and equations. Respective omegas for both test performances were 0.85 and

0.87. The students were told that their test performances would not impact their teacher-rated mathematics grade (i.e., a low-stakes test).

For a more comprehensive measure of students' mathematics achievement, teacher-rated mathematics grades at the end of fifth and sixth grades were retrieved from the school records. The Finnish grading scale in comprehensive education ranges from 4 (*fail*) to 10 (*excellent*) and is based on the Finnish National Core Curriculum. The criteria for the grade eighth are specified in the curriculum and teachers use them as a guideline for determining other grades. The focus of student assessment is on the students' learning and working skills. Although both formative and dialogical assessment practices and students' active role in these processes are emphasised in the curriculum, summative assessment has a strong position among Finnish teachers (e.g., Nieminen & Atjonen, 2023).

## Fidelity

Various indirect measures were taken to ensure and evaluate the fidelity of the implementation (e.g., Gresham et al., 2000), particularly in terms of adherence (i.e., the extent to which all activities were delivered as designed; Carroll et al., 2007). For the co-teachers, these measures included the above-mentioned workshops in the beginning of the school year, interviews on three occasions on implementation and progress, and a weekly check-list type diary. The diary included a description of the various principles of co-teaching: whether all aspects of teaching were shared, how much time was spent on planning and reflection, and an evaluation of the success of lessons on a scale from 1 (not at all satisfied) to 5 (very satisfied). The purpose of the evaluation was to evoke further reflection on what needs to be improved (e.g., Cook et al., 2017; Fluijt et al., 2016; Murawski & Lochner, 2018). To ensure the commitment of both schools and co-teachers, the special education teachers in the co-teaching dyads received monetary compensation from the municipality. This ensured that the special educators were not pulled to other tasks or to other classes from the co-taught mathematics classes.

Similar diaries and interviews were also completed by the solo teachers to evaluate whether they taught and planned mathematics by themselves (i.e., whether the aforementioned standard practice was maintained), or showed signs of reactivity to the experimental situation or compensatory rivalry (Shadish et al., 2002). The solo teachers' diary included weekly hours of special education teacher support in mathematics, whether this was in-class or pull-out, and if the special education teachers participated in the planning and evaluation processes.

According to the co-teacher interviews, the teachers planned, taught, and evaluated the mathematics lessons together, and this was corroborated by the co-teacher diaries. The diaries also revealed that mostly due to teacher absence (e.g., sick days and in-service training), co-teaching occurred in 70% of mathematics classes. However, the higher t-s ratio was still maintained as substitute teachers were used, although joint planning or assessment were not carried out during these times. Around three quarters of an hour per week was used for planning and reflection, and different co-teaching models such as team- and alternative teaching (see, Friend & Cook, 2013; Gardesten, 2023) were used flexibly. The co-teachers reported being satisfied with their co-teaching partnership, overall progress, and how the lessons were implemented throughout the year. These positive experiences were also reflected in the diaries: the perceived success of lessons was high ( $M=4.4$ ,  $SD=0.69$ ). The solo teacher dia-

ries and interviews revealed that the solo-taught classes were supported by mainly pull-out special education for approximately one hour a week, and the class teachers were responsible for planning and assessing students' progress.

Accordingly, our fidelity measures suggest that the design and implementation was realised as intended: the co-teachers planned, taught, and reflected as dyads, while the solo teachers planned and taught on their own, with some assistance from the special education teachers.

## Data analyses

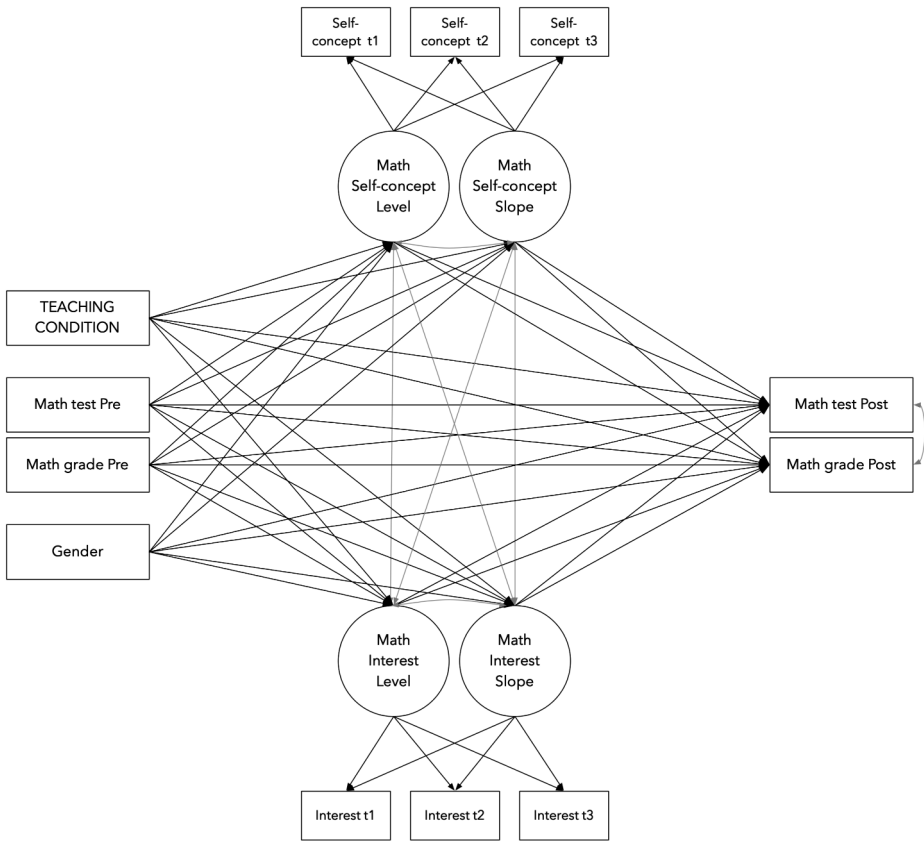
Latent growth curve modelling (LGCM) within the structural equation modelling framework was used, as it provides advantages in studying changes and development in multi-wave data (for an overview, see, Duncan & Duncan, 2009). In LGCM, observed variables are used to estimate latent factors that represent both the onset (initial level) and rate of change (slope) of the measured construct over time.

The analyses were carried out in five steps. First, to ensure that the measures reflect same constructs at each occasion, we conducted a stepwise procedure for testing longitudinal measurement invariance: configural (same number of factors and loading pattern over time), weak (identical factor loadings over time), strong (identical item intercepts over time), and strict (identical error variances over time) (Widaman & Reise, 1997). Second, separate unconditional univariate latent growth models for self-concept and individual interest were conducted to estimate trajectories over time. Third, an unconditional parallel process model was established to investigate the relations between the levels and changes of self-concept and interest. Fourth, the parallel process model was expanded to include predictors: teaching condition, test performance (at the beginning of sixth grade), teacher-rated mathematics grade (at end of fifth grade), and gender. Finally, the conditional model was extended to include outcome variables: test performance and teacher-rated mathematics grade at the end of sixth grade. The full model is illustrated in Fig. 2.

For evaluating the degree of measurement invariance, each model's fit was compared to the previous one by scaled difference  $\chi^2$ -tests (Satorra & Bentler, 2010), and by criteria following the recommendations by Chen (2007): for testing weak invariance, a change of  $\leq -0.005$  in CFI, supplemented with a change in RMSEA  $\geq 0.010$ , or a change in SRMR  $\geq 0.025$  would indicate noninvariance, and for strong and strict models a change of  $\geq -0.005$  in CFI, and a change  $\geq 0.010$  in RMSEA or a change of  $\geq 0.005$  in SRMR would indicate noninvariance. As to the latent growth curve models, model fit was evaluated using Comparative Fit Index (CFI: cutoff value  $> 0.95$ ), Root Mean Square Error of Approximation (RMSEA: cutoff value  $< 0.06$ ), and the Standardised Root Mean Square Residual (SRMR: cutoff value  $< 0.08$ ) (Hu & Bentler, 1999).

IBM SPSS Statistics v27 and Jamovi 2.3.18 software (the jamovi project, 2022) were used for descriptive statistics, while all other analyses were conducted using Mplus statistical software 8.6 (Muthén & Muthén, 2019). We used maximum likelihood estimation with robust standard errors (MLR) for all models<sup>2</sup>, and missing data (ranging from 0 to 9% across

<sup>2</sup> Despite the small number of classes, we calculated intraclass correlations (ICCs) to address the extent of possible clustering within the nested data. The ICCs for the Time 1 measures of self-concept, interest, test performance, and teacher-rated mathematics grade were 0.044, 0.069, 0.109, and 0.111, respectively, with corresponding design effects  $(1 + (m - 1) \times \text{ICC})$ ; where  $m$  is the average number of subjects per class; see



**Fig. 2** Hypothetical model with predictors and outcomes

all included variables; see Table 1) were handled with full-information maximum likelihood estimation (Little's MCAR test:  $\chi^2(53)=59.71$ ;  $p=.245$ ).

## Results

### Descriptive statistics and measurement invariance

Descriptive statistics and bivariate correlations are shown in Table 1. Both self-concept and interest demonstrated high rank-order stability over time, with between-measurement correlations ranging from 0.80 to 0.85 for self-concept and 0.78 to 0.83 for interest. Similarly, both teacher-rated mathematics achievement ( $r=.82$ ,  $p<.001$ ) and test performance ( $r=.79$ ,

Muthén & Satorra, 1995) of 1.87, 2.37, 3.20, and 3.16, indicating minor clustering in the performance measures. In Mplus, less biased standard errors can be estimated for nested data using the TYPE=COMPLEX option, but this generally requires a much larger number of groups in the data (Maas & Hox, 2005). However, for the sake of comparison, we ran a parallel full model using this option. These results and a corresponding discussion are included in the Supplementary.

**Table 2** Tests of longitudinal measurement invariance

Model	$\chi^2$ MLR	df	CFI	RMSEA	SRMR	$\Delta$ df	$p(\Delta\chi^2)$	$\Delta$ CFI	$\Delta$ RMSEA	$\Delta$ SRMR
Configural	276.379***	213	0.982	0.045	0.036					
Weak	289.826***	225	0.982	0.044	0.045	12	0.360	0.000	-0.001	0.009
Strong	302.951***	237	0.981	0.044	0.046	12	0.368	-0.001	0.000	0.001
Strict	328.369***	253	0.979	0.045	0.044	16	0.086	-0.002	0.001	-0.002

Note. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

$p < .001$ ) showed high intraindividual stability over time. The relatively high intercorrelations between teacher-rated grades and test performance ( $r = .61 - 0.67$ ,  $p < .001$ ) demonstrated a strong link between the different indicators of mathematics performance. Gender was somewhat associated with earlier ( $r = -.21$ ,  $p = .012$ ) and later ( $r = -.15$ ,  $p = .064$ ) mathematics achievement, showing girls to have slightly higher teacher-rated grades at the end of fifth and sixth grade, but not with test performance. The stepwise tests of longitudinal measurement invariance (Table 2) showed strict invariance in both self-concept and interest, thus permitting valid inferences of changes in means over time.

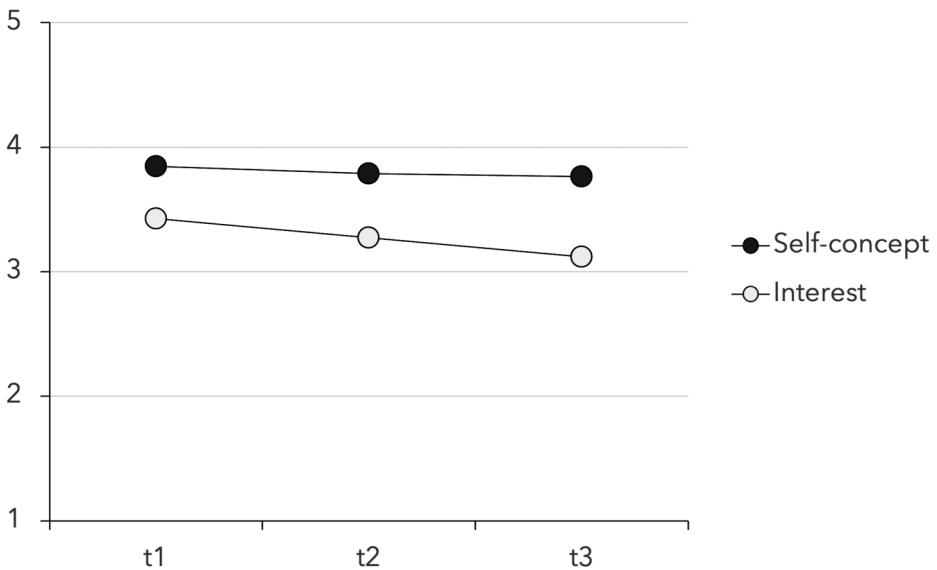
### Mean-level changes in mathematics self-concept and interest

To address the first research question, separate univariate latent growth models for self-concept and interest were estimated. The linear model for self-concept had a satisfactory fit,  $\chi^2(1) = 3.70$ ,  $p = .055$ , CFI = 0.992, RMSEA = 0.136 (90% CI: 0.000–0.295), SRMR = 0.015. However, given the patterning of observed means (Table 1), we also estimated a latent basis model, where the first and third loadings were fixed to zero and one, respectively, and the second loading was estimated freely (Grimm et al., 2011; Wang & Wang, 2020). Residual variances were fixed to be equal. Since this model fit the data better,  $\chi^2(2) = 2.62$ ,  $p = .270$ , CFI = 0.998, RMSEA = 0.046 (90% CI: 0.000–0.177), SRMR = 0.027, we chose it for the subsequent analyses. The results from this model showed a negative slope in students' self-concept ( $M = -0.081$ ,  $p = .077$ ), thus indicating a small overall decline in mathematics self-concept over time (see Fig. 3). Variances of both the onset ( $S^2 = 0.470$ ,  $p < .001$ ) and the slope ( $S^2 = 0.141$ ,  $p = .020$ ) were statistically significant, hence demonstrating significant individual variability in the initial levels and changes over time in students' mathematics self-concept.

The univariate linear model for interest fitted the data well,  $\chi^2(1) = 0.20$ ,  $p = .653$ , CFI = 1.000, RMSEA = 0.000 (90% CI: 0.000–0.168), SRMR = 0.005. Similar to self-concept, the results showed an overall decrease in individual interest ( $M = -0.154$ ,  $p < .001$ ) as well as significant individual differences in the variance of both the onset ( $S^2 = 0.938$ ,  $p < .001$ ) and change over time ( $S^2 = 0.135$ ,  $p = .017$ ).

### Concurrent and longitudinal relationships between self-concept and interest

To investigate how the levels and changes of self-concept and interest were mutually related, an unconditional parallel process model was estimated. The model had an excellent fit,  $\chi^2(8) = 10.14$ ,  $p = .256$ , CFI = 0.996, RMSEA = 0.043 (90% CI: 0.000–0.112), SRMR = 0.025. An inspection of latent correlations (Table 3) revealed both the initial levels ( $r = .65$ ,  $p < .001$ ) and slopes ( $r = .64$ ,  $p < .001$ ) of self-concept and interest to be highly



**Fig. 3** Model estimated means of the developmental trajectories of mathematics self-concept and interest

**Table 3** Descriptive statistics and latent correlations for the initial levels and changes in self-concept and interest

	M	s.e.	Self-concept: Level	Self-concept: Slope	Interest: Level	Interest: Slope
Self-concept: Level	3.84***	0.06	1.00			
Self-concept: Slope	-0.08†	0.05	-0.07	1.00		
Interest: Level	3.43***	0.09	0.65***	-0.20	1.00	
Interest: Slope	-0.16***	0.04	-0.13	0.64***	-0.22†	1.00

Note. †  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

correlated. In other words, students' mathematics self-concept and interest were strongly connected, both concurrently and longitudinally: when students' self-concept became less positive, interest declined as well, and vice versa.

### Predictions of the levels and changes in self-concept and interest

Next, predictors (i.e., teaching condition, test performance, teacher-rated grade, and gender) were added to the parallel process model (Table 4). This conditional model fit the data well,  $\chi^2(16)=19.27$ ,  $p=.255$ , CFI=0.995, RMSEA=0.038 (90% CI: 0.000–0.091), SRMR=0.023. Only teacher-rated grade predicted positively the onset of self-concept ( $\beta = 0.61$ ,  $p < .001$ ). That is, the students who had higher teacher-rated grades at the end of fifth grade reported higher mathematics self-concept in the beginning of sixth grade. Gender had a small effect on the onset of self-concept ( $\beta = 0.27$ ,  $p=.064$ ), meaning that boys reported somewhat higher mathematics self-concept at the beginning of sixth grade. Contrary to our expectation, teaching condition was not predictive of changes in self-concept and interest.

**Table 4** Standardised effects from the full model

Predictor	Self-concept: Level			Self-concept: Slope			Interest: Level			Interest: Slope			Math test Post			Math grade Post		
	$\beta$	Z	p	$\beta$	Z	p	$\beta$	Z	p	$\beta$	Z	p	$\beta$	Z	p	$\beta$	Z	p
Teaching condition	0.06	0.42	0.675	0.14	0.54	0.590	0.01	0.04	0.967	0.10	0.48	0.629	-0.00	-0.03	0.977	0.25	2.64	0.008
Math test Pre	0.11	1.31	0.191	0.04	0.24	0.814	0.14	1.19	0.234	-0.09	-0.78	0.434	0.61	9.06	<0.001	0.11	1.67	0.094
Math grade Pre	0.61	7.68	<0.001	0.18	1.05	0.292	0.20	1.65	0.100	0.12	0.88	0.380	0.16	1.40	0.161	0.62	7.04	<0.001
Gender	0.27	1.86	0.063	-0.05	-0.17	0.867	-0.01	-0.03	0.976	-0.18	-0.88	0.378	-0.04	-0.39	0.694	-0.02	-0.22	0.824
Self-concept: Level													0.14	1.10	0.271	0.17	1.33	0.184
Self-concept: Slope													-0.01	-0.12	0.906	0.16	1.29	0.196
Interest: Level													-0.04	-0.49	0.625	0.00	0.003	0.997
Interest: Slope													0.04	0.40	0.691	-0.11	-1.04	0.297

Note. Gender coding: 0 = girls, 1 = boys; Group coding: 0 = solo teaching, 1 = co-teaching

## Predictions of the outcomes

In the final step, the conditional model was expanded to include our two outcomes: teacher-rated mathematics grade and mathematics test performance at the end of sixth grade. The full model had an excellent fit,  $\chi^2(20)=20.30$ ,  $p=.440$ , CFI=1.000, RMSEA=0.010 (90% CI: 0.000–0.073), SRMR=0.019. Interestingly, the teaching condition predicted teacher-rated grade ( $\beta = 0.25$ ,  $p=.008$ ), over and above the significant effects of both previous teacher-rated grade ( $\beta = 0.62$ ,  $p < .001$ ) and test performance ( $\beta = 0.11$ ,  $p=.094$ ), but, against our assumptions, neither the initial levels nor changes in self-concept and interest predicted later mathematics performance. All effects are reported in Table 4<sup>3</sup>.

## Discussion

The purpose of this study was to investigate the role of co-teaching in the development of students' motivation and achievement in the domain of mathematics. More specifically, we investigated how sixth graders' mathematics self-concept and individual interest changed over the final year of primary school, how these changes were connected with each other and mathematics achievement, and, most importantly, whether these changes differed between solo and co-teaching groups.

Regarding changes in motivation, students' mathematics self-concept and interest declined over time, which concurs with previous findings (Denissen et al., 2007; Frenzel et al., 2010; Jacobs et al., 2002; Pinxten et al., 2014; Savolainen et al., 2018). More interestingly, those changes were also highly correlated (Petersen & Hyde, 2017), meaning that the decrease in self-concept was associated with a parallel decrease in interest, and vice versa. This, in a sense, supports the view of cyclically developing relations between mathematics self-concept and interest (Marsh et al., 2005; Petersen & Hyde, 2017), although our data do not permit any inferences about their causal predominance.

Whilst our findings concerning the decline in mathematics motivation concur with previous research, it is not clear why this occurred, particularly within just one school year. It has been postulated that such a negative change may stem from normative changes associated with adolescence, the nature of mathematics as a school subject both in terms of content (e.g., cumulative learning and topics becoming increasingly more complex) and pedagogical practice (e.g., repetitive tasks with a focus on correctness), or due to more general contextual changes such as decrease in students' sense of autonomy and relatedness (Frenzel et al., 2010; Ryan & Deci, 2017; Schiefele, 2009; Stipek, 1996; Stipek & Iver, 1989). Given that we found both overall decline and rank-order stability in mathematics self-concept and interest over time, such more systemic reasons for the changes would seem most likely. Considering that an extensive body of research has found both self-concept and interest to be consequential in learning, more attention should be paid to effective instructional practices that would counter the negative changes. In this context, it is also important to note that

<sup>3</sup> To evaluate the generalizability of our model and to address concerns about overfitting due to complexity, we ran additional models omitting gender from both the conditional,  $\chi^2(14)=18.738$ ,  $p=.175$ , CFI=0.993, RMSEA=0.049 (90% CI: 0.000–0.101), SRMR=0.024, and full model,  $\chi^2(18)=19.81$ ,  $p=.344$ , CFI=0.998, RMSEA=0.027 (90% CI: 0.000–0.082), SRMR=0.020. Since the model fit and parameter estimates were virtually no different from the reported models, we considered them appropriate.



the COVID–19 pandemic might have played a role here. Although the last measurement took place shortly after the transition to online teaching, the concerns about the impact of the COVID–19 pandemic on schools started earlier. Despite this, however, the consistency of the observed changes suggests that the pandemic did not necessarily amplify the decline.

Despite the concurrent connections observed between teacher-rated grades and self-concept and interest that also accord with prior research (Marsh et al., 2005; Petersen & Hyde, 2017), the changes in self-concept and interest were independent of previous achievement. This contrasts with our expectations, as well as with the limited findings available reporting a link between better mathematics achievement and less steep negative changes in motivation (Denissen et al., 2007). Consequently, higher achievement at the beginning of sixth grade did not seem to guarantee sustained motivation in mathematics across the school year. This independence from preceding achievement may also point to more systemic reasons for the unfavourable changes in self-concept and interest, which implies that motivational support would be relevant for all students regardless of their skill-level. Perhaps following students' classroom experiences more intensively (e.g., through experience sampling methodology) would reveal more about the sources of students' declining motivation in mathematics, including their perceptions of the classroom climate and the subject itself (e.g., Neubauer et al., 2022; Talić et al., 2022).

Regarding the main theme, the role of co-teaching, we found it to be no more successful in supporting students' declining motivation in mathematics than solo teaching. This is somewhat disappointing, given the previous findings showing both that teachers can be effective in reinforcing students' self-concept (O'Mara et al., 2006) and interest (Kiemer et al., 2015; Renninger & Hidi, 2011; Rotgans & Schmidt, 2017), and that indirect interventions aiming to improve student-teacher interaction and pedagogical practices – such as this study – may well work (Marsh & Craven, 1997; O'Mara et al., 2006; Watson et al., 2019).

Interestingly, however, we found teaching condition to predict teacher-rated grades at the end of the school year, although not through changes in self-concept and interest as hypothesised. That is, compared to students in the solo teaching group, the mathematics grades of students in the co-teaching group improved more. This might be a valid outcome due to changes in either teaching practices or assessment, or an artefact due to experimental bias. It indeed is possible that students' gain in achievement was true due to the efforts of co-teaching (e.g., Jones & Winters, 2022), but just not mediated by motivation, or reflected in test performance. Perhaps the students were more cognitively engaged in the co-teaching condition (Blatchford et al., 2011; Finn et al., 2003; Lochner et al., 2019), which, in turn, translated into better achievement. That these efforts were not similarly seen in test performance might be because the tests measured mostly basic mathematics skills, while teacher evaluations likely reflected a more comprehensive view into students' mathematics achievement (see, Brookhart et al., 2016). Including a broader set of mathematics tests might thus be something to consider in future studies.

It is also possible that the relatively higher gain in co-taught students' achievement was a consequence of improved assessment practices. The higher teacher-student ratio and the opportunity to reflect on student learning with a teaching partner might have actualised in more frequent and thorough student assessments, including the use of individual frame of reference standard, as opposed to a focus on end-of-period summative evaluation. That is, perhaps students' improvement in the co-taught group was due to continuous evaluation of their progress that provided the students with more accurate feedback.

However, we cannot exclude the possibility that the said gain was due to a bias in assessment. Although teacher-ratings have been consistently linked with test performance and other measures of achievement, they can also be prone to subjectivity and bias, as such assessments represent various cognitive and non-cognitive factors valued by teachers themselves (Brookhart et al., 2016; Lekholm Klapp & Cliffordson, 2009). Having another teacher in class may have increased this subjectivity even further. It is thus possible that the co-teachers' positive experiences (as revealed in their diaries) of co-teaching along with an experimental effect introduced some additional bias (e.g., confirmation bias) to their grading. Clearly more research is needed to address the validity of the observed gains as well as the possible sources of bias.

Given the above, why did co-teaching fail to counter the negative changes in students' mathematics motivation? One possibility is that the co-teaching practices were simply not pertinent to students' self-concept and interest, but rather promoted student engagement. It may also be that a single school year might be too short a period to have a meaningful impact on either student self-concept or interest through indirect influence. Together, these findings suggest that it might be beneficial to incorporate into co-teaching practices explicit instructional practices proven to support student motivation in mathematics (e.g., differentiating, context personalisation, and feedback) or even interventions directly targeted at self-concept and interest.

All the above also exemplifies that the nature of our study could be considered both as a strength and a weakness. While being ecologically more valid than a highly controlled experiment (Bronfenbrenner, 1977), the everyday events taking place in the classroom occurred naturally without external influence or interference. Thus, despite our fidelity measures indicating successful implementation, we had no means of ensuring what exactly happened in the classrooms, in both solo and co-teaching groups, or the extent to which the learning and teaching activities taking place over the year were influenced by factors other than we intended or were aware of. We also acknowledge that our relatively small sample size (which was mainly due to practical realities) might have limited the statistical power to detect small effects.

Nevertheless, we would consider the outcomes of this rather laborious endeavour promising from the perspective of practice, and informative from the perspective of research. Our experiences are not in disagreement with previous inferences stating that the implementation of co-teaching is resource-intensive (Cook et al., 2017) and far from straightforward (e.g., Friend, 2008; Pratt, 2014), and there is much to be learned from the present efforts, but we do consider the findings encouraging. Naturally, there is room for improvement, and many issues remain open, but in conclusion, we view co-teaching as having the potential for providing an auspicious approach to developing and supporting not only students' learning but also teaching itself.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11251-024-09678-4>.

**Funding** Open access funding provided by University of Eastern Finland (including Kuopio University Hospital). This study was supported by a grant from the Finnish Cultural Foundation, South Karelia Regional fund (grant #05202282) to the first author.

## Declarations

**Competing interests** The authors report there are no competing interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94(3), 545–561. <https://doi.org/10.1037/0022-0663.94.3.545>
- Arens, A., & Hasselhorn, M. (2015). Differentiation of competence and affect self-perceptions in elementary school students: Extending empirical evidence. *European Journal of Psychology of Education; a Journal of Education and Development*, 30(4), 405–419. <https://doi.org/10.1007/s10212-015-0247-8>
- Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, 103(4), 970. <https://doi.org/10.1037/a0025047>
- Bernacki, M. L., & Walkington, C. (2018). The role of situational interest in personalized learning. *Journal of Educational Psychology*, 110(6), 864–881. <https://doi.org/10.1037/edu0000250>
- Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, 21(6), 715–730. <https://doi.org/10.1016/j.learninstruc.2011.04.001>
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40. <https://doi.org/10.1023/A:1021302408382>
- Bright, B. (1996). Reflecting on ‘Reflective practice’. *Studies in the Education of Adults*, 28(2), 162–184. <https://doi.org/10.1080/02660830.1996.11730638>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *The American Psychologist*, 32(7), 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75(5), 631–661. <https://doi.org/10.1037/0022-0663.75.5.631>
- Brophy, J. (2011). Motivating students in classrooms. In S. Jarvela (Ed.), *Social and emotional aspects of learning* (pp. 50–56). Elsevier.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 1–9. <https://doi.org/10.1186/1748-5908-2-40>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Conderman, G., & Hedin, L. (2015). Differentiating instruction in co-taught classrooms for students with emotional/behavioural difficulties. *Emotional and Behavioural Difficulties*, 20(4), 349–361. <https://doi.org/10.1080/13632752.2014.976918>
- Cook, S. C., Meduffie-Landrum, K. A., Oshita, L., & Cook, B. G. (2017). Co-teaching for students with disabilities. In J. Kauffman, D. Hallahan & P. Pullen (Eds.). *Handbook of special education* (2nd ed., pp. 233–248). Routledge.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A Meta-Analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>

- Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know i am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, 78(2), 430–447. <https://doi.org/10.1111/j.1467-8624.2007.01007.x>
- Dewey, J. (1913). *Interest and effort in education*. Riverside.
- Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass; Soc Personal Psychol Compass*, 3(6), 979–991. <https://doi.org/10.1111/j.1751-9004.2009.00224.x>
- Durik, A. M., Shechter, O. G., Noh, M., Rozek, C. S., & Harackiewicz, J. M. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion*, 39(1), 104–118. <https://doi.org/10.1007/s11031-014-9419-0>
- Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The why's of class size: Student behavior in small classes. *Review of Educational Research*, 73(3), 321–368. <https://doi.org/10.3102/00346543073003321>
- Finnish National Board on Research Integrity TENK (2019). *The ethical principles of research with human participants and ethical review in the human sciences in Finland*[https://tenk.fi/sites/default/files/2021-01/Ethical\\_review\\_in\\_human\\_sciences\\_2020.pdf](https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf)
- Fluijt, D., Bakker, C., & Struyf, E. (2016). Team-reflection: The missing link in co-teaching teams. *European Journal of Special Needs Education*, 31(2), 187–201. <https://doi.org/10.1080/08856257.2015.1125690>
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507–537. <https://doi.org/10.1111/j.1532-7795.2010.00645.x>
- Friend, M. (2008). Co-teaching: A simple solution that isn't simple after all. *Journal of Curriculum and Instruction*, 2(2), 9–19. <https://doi.org/10.3776/joci.2008.v2n2p>
- Friend, M., & Cook, L. (2013). *Interactions: Collaboration skills for school professionals* (7th ed.). Pearson.
- Friend, M., Cook, L., Hurlley-Chamberlain, D., & Shamberger, C. (2010). Co-teaching: An illustration of the complexity of collaboration in special education. *Journal of Educational and Psychological Consultation: Collaboration in Special Education*, 20(1), 9–27. <https://doi.org/10.1080/10474410903535380>
- Ganley, C. M., & Lubienski, S. T. (2016). Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, 47, 182–193. <https://doi.org/10.1016/j.lindif.2016.01.002>
- Gardesten, M. (2023). How co-teaching may contribute to inclusion in mathematics education: A systematic literature review. *Education Sciences*, 13(7), 677. <https://doi.org/10.3390/educsci13070677>
- Gniewosz, B., Eccles, J. S., & Noack, P. (2012). Secondary school transition and the use of different sources of information for the construction of the academic self-concept. *Social Development*, 21(3), 537–557. <https://doi.org/10.1111/j.1467-9507.2011.00635.x>
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, 15(4), 198–205. [https://doi.org/10.1207/SLDRP1504\\_4](https://doi.org/10.1207/SLDRP1504_4)
- Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, 82(5), 1357–1371. <https://doi.org/10.1111/j.1467-8624.2011.01630.x>
- Hakkarainen, A. M., Holopainen, L. K., & Savolainen, H. K. (2016). The impact of learning difficulties and socioemotional and behavioural problems on transition to postsecondary education or work life in Finland: A five-year follow-up study. *European Journal of Special Needs Education*, 31(2), 171–186. <https://doi.org/10.1080/08856257.2015.1125688>
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in Education. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 220–227. <https://doi.org/10.1177/2372732216655542>
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60(4), 549–571. <https://doi.org/10.3102/00346543060004549>
- Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review*, 1(2), 69–82. <https://doi.org/10.1016/j.edurev.2006.09.001>
- Hidi, S., & Ainley, M. (2002). Interest and adolescence. In F. Pajares, & T. C. Urdan (Eds.), *Academic motivation of adolescents* (pp. 247–275). Information Age Publishing Inc.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. [https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4)
- Høgheim, S., & Reber, R. (2015). Supporting interest of middle school students in mathematics through context personalization and example choice. *Contemporary Educational Psychology*, 42, 17–25. <https://doi.org/10.1016/j.cedpsych.2015.03.006>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527. <https://doi.org/10.1111/1467-8624.00421>
- Jones, N., & Winters, M. A. (2022). Are two teachers Better Than one? The effect of co-teaching on students with and without disabilities. *Journal of Human Resources*, 57(4), 1–37. <https://doi.org/10.3368/jhr.0420-10834R3>
- Kaplan, A., & Patrick, H. (2016). Learning environments and motivation. In K. Wentzel, & D. Miele (Eds.), *Handbook of motivation at school* (pp. 251–274). Routledge.
- Katz, I., Assor, A., Kanat-Maymon, Y., & Bereby-Meyer, Y. (2006). Interest as a motivational resource: Feedback and gender matter, but interest makes the difference. *Social Psychology of Education*, 9(1), 27–42. <https://doi.org/10.1007/s11218-005-2863-7>
- Kiemer, K., Gröschner, A., Pehmer, A. K., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction*, 35, 94–103. <https://doi.org/10.1016/j.learninstruc.2014.10.003>
- Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32(5), 448–470. <https://doi.org/10.2307/749801>
- Lazarides, R., Gaspard, H., & Dicke, A. (2019). Dynamics of classroom motivation: Teacher enthusiasm and the development of math interest and teacher support. *Learning and Instruction*, 60, 126–137. <https://doi.org/10.1016/j.learninstruc.2018.01.012>
- Lekholm Klapp, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*, 15(1), 1–23. <https://doi.org/10.1080/13803610802470425>
- Lochner, W. W., Murawski, W. W., & Daley, J. T. (2019). The effect of co-teaching on student cognitive engagement. *Theory & Practice in Rural Education*, 9(2), 6–19. <https://doi.org/10.3776/tpre.2019.v9n2p6-19>
- Lütke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology*, 30(3), 263–285. <https://doi.org/10.1016/j.cedpsych.2004.10.002>
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92. <https://doi.org/10.1027/1614-2241.1.3.86>
- Marsh, H. (1990). *SDQ 1 manual & research monograph: Self-description questionnaire*. Psychological Corporation & Harcourt Brace Jovanovich.
- Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Academic.
- Marsh, H. W., & Craven, R. (2000). Swimming in the school: Expanding the scope of the big fish 75 little pond effect. In R. Craven, & H. Marsh (Eds.), *Self-concept theory, research and practice: Advances for the new millennium* (pp. 419–425). 2000 SELF Research Centre.
- Marsh, H. W., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2), 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>
- Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81(1), 59–77. <https://doi.org/10.1348/000709910X503501>
- Marsh, H., Trautwein, U., Lütke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416. <https://doi.org/10.1111/j.1467-8624.2005.00853.x>
- Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in Mathematics: Findings, generalizations, and criticisms of the Research. *Journal for Research in Mathematics Education*, 30(1), 65–88. <https://doi.org/10.2307/749630>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/international-results/>
- Murawski, W. W., & Lochner, W. (2011). Observing co-teaching: What to ask for, look for, and listen for. *Intervention in School and Clinic*, 46(3), 174–183. <https://doi.org/10.1177/1053451210378165>
- Murawski, W., & Lochner, W. (2018). *Beyond co-teaching basics: A data-driven, no-fail model for continuous improvement*. ASCD.
- Murawski, W. W., & Swanson, L. H. (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>
- Muthén, L. K., & Muthén, B. (2019). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267. <https://doi.org/10.2307/271070>

- Neubauer, A. B., Schmidt, A., Schmiedek, F., & Dirk, J. (2022). Dynamic reciprocal relations of achievement goals with daily experiences of academic success and failure: An ambulatory assessment study. *Learning and Instruction, 81*, 101617. <https://doi.org/10.1016/j.learninstruc.2022.101617>
- Nieminen, J. H., & Atjonen, P. (2023). The assessment culture of mathematics in Finland: A student perspective. *Research in Mathematics Education, 25*(2), 243–262. <https://doi.org/10.1080/14794802.2022.2045626>
- Nuthall, G. (2005). The cultural myths and realities of classroom teaching and learning: A personal journey. *Teachers College Record (1970), 107*(5), 895–934. <https://doi.org/10.1111/j.1467-9620.2005.00498.x>
- Nuutila, K., Tapola, A., Tuominen, H., Kupiainen, S., Pásztor, A., & Niemivirta, M. (2020). Reciprocal predictions between interest, self-efficacy, and performance during a task. *Frontiers in Education, 5*, 36. <https://doi.org/10.3389/educ.2020.00036>
- O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*(3), 181–206. [https://doi.org/10.1207/s15326985ep4103\\_4](https://doi.org/10.1207/s15326985ep4103_4)
- OECD. (2013). PISA 2012 Assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. *OECD Publishing*. <https://doi.org/10.1787/9789264190511-en>
- Official Statistics of Finland (OSF) (2021, June 8). *Support for learning. Statistics Finland*. Retrieved January 10, 2021, from [http://www.stat.fi/til/erop/2020/erop\\_2020\\_2021-06-08\\_tie\\_001\\_en.html](http://www.stat.fi/til/erop/2020/erop_2020_2021-06-08_tie_001_en.html)
- Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* Institute of Education, University of London. National Research and Development Centre for Adult Literacy and Numeracy.
- Pekrun, R. (2016). Academic emotions. In K. Wentzel, & D. Miele (Eds.), *Handbook of motivation at school* (pp. 120–144). Routledge.
- Petersen, J. L., & Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology, 37*(4), 438–456. <https://doi.org/10.1080/01443410.2015.1076765>
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology, 84*(1), 152–174. <https://doi.org/10.1111/bjep.12028>
- Pratt, S. (2014). Achieving symbiosis: Working through challenges found in co-teaching to achieve effective co-teaching relationships. *Teaching and Teacher Education, 41*, 1–12. <https://doi.org/10.1016/j.tate.2014.02.006>
- Räsänen, P. (2004). *RMAT-laskutaidon testi 9–12-vuotiaille: Käsikirja*. Niilo Mäki Instituutti.
- Remler, D. K., & Van Ryzin, G. G. (2011). *Research methods in practice: Strategies for description and causation*. SAGE.
- Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist, 46*(3), 168–184. <https://doi.org/10.1080/00461520.2011.587723>
- Rotgans, J. I., & Schmidt, H. G. (2017). Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educational Psychology, 49*, 175–184. <https://doi.org/10.1016/j.cedpsych.2017.02.003>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.
- Rytivaara, A., & Kershner, R. (2012). Co-teaching as a context for teachers' professional learning and joint knowledge construction. *Teaching and Teacher Education, 28*(7), 999–1008. <https://doi.org/10.1016/j.tate.2012.05.006>
- Saloviita, T., & Takala, M. (2010). Frequency of co-teaching in different teacher categories. *European Journal of Special Needs Education, 25*(4), 389–396. <https://doi.org/10.1080/08856257.2010.513546>
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*, 243–248. <https://doi.org/10.1007/s11336-009-9135-y>
- Savolainen, P. A., Timmermans, A. C., & Savolainen, H. K. (2018). Part-time special education predicts students' reading self-concept development. *Learning and Individual Differences, 68*, 85–95. <https://doi.org/10.1016/j.lindif.2018.10.005>
- Schiefele, U. (2009). Situational and individual interest. In K. Wentzel, & D. Miele (Eds.), *Handbook of motivation at school* (pp. 197–222). Routledge.
- Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children, 73*(4), 392–416. <https://doi.org/10.1177/001440290707300401>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. <https://doi.org/10.3102/00346543046003407>



- Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist*, 37(4), 233–244. [https://doi.org/10.1207/S15326985EP3704\\_3](https://doi.org/10.1207/S15326985EP3704_3)
- Stipek, D. J. (1996). Motivation and instruction. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 85–113). Prentice Hall International.
- Stipek, D. J., & Iver, D. M. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, 60(3), 521. <https://doi.org/10.2307/1130719>
- 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>
- Sweigart, C. A., & Landrum, T. J. (2015). The impact of number of adults on instruction: Implications for co-teaching. *Preventing School Failure: Alternative Education for Children and Youth*, 59(1), 22–29. <https://doi.org/10.1080/1045988X.2014.919139>
- Talić, I., Scherer, R., Marsh, H. W., Greiff, S., Möller, J., & Niepel, C. (2022). Uncovering everyday dynamics in students' perceptions of instructional quality with experience sampling. *Learning and Instruction*, 81, 101594. <https://doi.org/10.1016/j.learninstruc.2022.101594>
- Tapola, A., Veermans, M., & Niemivirta, M. (2013). Predictors and outcomes of situational interest during a science learning task. *Instructional Science*, 41(6), 1047–1064. <https://doi.org/10.1007/s11251-013-9273-6>
- The jamovi project (2022). *jamovi* (Version 2.3) [Computer Software]. <https://www.jamovi.org>
- Tsai, Y., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. M. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, 100(2), 460–472. <https://doi.org/10.1037/0022-0663.100.2.460>
- Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian Journal of Educational Research*, 58(6), 734–756. <https://doi.org/10.1080/00313831.2014.904419>
- Villa, R. A., Thousand, J. S., & Nevin, A. (2008). *A guide to co-teaching: Practical tips for facilitating student learning*. Corwin.
- Wang, J., & Wang, X. (2020). *Structural equation modeling: applications using Mplus* (Second edition). Wiley.
- Watson, P. W. S. J., Rubie-Davies, C. M., & Meissel, K. (2019). Mathematics self-concept in New Zealand elementary school students: Evaluating age-related decline. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02307>
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association. <https://doi.org/10.1037/10222-009>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Wilson, G. L., & Michaels, C. A. (2006). General and Special Education Students' perceptions of co-teaching: Implications for secondary-level literacy instruction. *Reading & Writing Quarterly*, 22(3), 205–225. <https://doi.org/10.1080/10573560500455695>