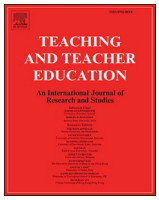
Teaching and Teacher Education 153 (2025) 104847

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate



Research paper

Influence of track placement and teachers' perceptions of children's academic schoolwork skills on the development of children's motivational self-beliefs and achievement

Satu Koivuhovi [a,*], Alexander Jung [b], Elina Kilpi-Jakonen [a], Todd D. Little [c], Mari-Pauliina Vainikainen [d]

[a] INVEST (Inequalities, Interventions and a New Welfare State), University of Turku, Finland
[b] Universität Tübingen, Hector-Institut für Empirische Bildungsforschung Europastraße 6, 72072, Tubingen, Germany
[c] Texas Tech University, College of Education, 3002 18th Street Lubbock, TX 7409, USA
[d] Faculty of Education and Culture, Åkerlundinkatu 5, 33014, Tampere University, Finland

ARTICLE INFO

Keywords:
Motivation
Teacher's perceptions
Tracking effects
Longitudinal study
Achievement
Elementary school

ABSTRACT

In this study, we analyzed longitudinal data from 1065 Finnish elementary school students to examine how teachers' perceptions of students' academic schoolwork skills and track placement influence children's motivational self-beliefs and achievement. Our findings indicate, that teachers' perceptions of students, which varied notably based on the students' gender, their mothers' education level, and their class types, significantly affected children's motivation and achievement. Teachers viewed girls, children of more educated mothers, and students in classes with a special emphasis more positively. Moreover, studying in a class with a special emphasis was associated with more positive developments in achievement and ability beliefs.

1. Introduction

This study examines the development of children's motivational self-beliefs and achievement during elementary school years and the role of track placement and teachers' perceptions of children's schoolwork skills in it.

Children's motivational self-beliefs develop reciprocally with experiences of achievement (Guay et al., 2003; Pajares & Schunk, 2001). At school, children regularly encounter achievement situations and receive feedback on their performance, shaping their perceptions of themselves as learners. Teachers play a crucial role in providing competence-related feedback, offering valuable insights into students' academic performance (Skaalvik & Hagtvet, 1990). Previous research has demonstrated that teachers' beliefs and perceptions of students' abilities and motivation, alongside their expectations, significantly influence students' self-perceptions as learners (e.g., Jussim et al., 2009; Pesu et al., 2016; Upadyaya & Eccles, 2014). As teachers' expectations and perceptions have been found to be influenced by the grouping of students (e.g., Andersen, 2018; Bohlmann & Weinstein, 2013), they are proposed as one of the mechanisms through which tracking influences individual student's learning (Pallas et al., 1994).

Tracking[1] students based on their abilities or other aptitudes is common practice in many countries. Various allocation and grouping practices can be seen as an attempt to manage the diversity and individual differences in education and a way of optimizing teaching to maximize benefits for all students (Dupriez et al., 2008). Although the effects of tracking have been studied extensively, the question of tracking remains an ongoing debate in educational research (e.g., Steenbergen-Hu et al., 2016; Zimmer, 2003). Most of the studies analyzing tracking effects in relation to learning outcomes have focused on academic achievement (e.g., Chiu et al., 2017; Collins & Gan, 2013). Less research has been conducted in the field of motivation, even though motivation influences not only academic achievement, but also social and emotional development and well-being (Elliot et al., 2017).

Gamoran (2010) has suggested that instructional variation across tracks (i.e., differing curricula) could be one of the main reasons

* Corresponding author.
E-mail address: satu.koivuhovi@utu.fi (S. Koivuhovi).

[1] Use of different concepts referring to allocation or grouping practices seems somewhat inconsistent in research (see, Koivuhovi, 2021). In this paper, we use the term *tracking* as a general concept referring to various allocation and grouping practices, in which students are sorted into groups based on their abilities or other aptitudes.

https://doi.org/10.1016/j.tate.2024.104847
Received 22 April 2024; Received in revised form 25 September 2024; Accepted 2 November 2024
Available online 8 November 2024

explaining differences in achievement between tracks. However, examining tracking effects is challenging, due to influences from different sources being confounded and difficult to separate in statistical analyses. In the present investigation, we have the opportunity to circumvent this challenge by utilizing longitudinal data from 1065 Finnish elementary school students. Although the Finnish basic education system is often regarded as a model example of a non-tracked system where all children adhere to the same national core curriculum, it has been less emphasized that a nuanced form of tracking exists (Seppänen et al., 2023). Introduced in the 1990s, "classes with a special emphasis"—which provide additional instruction in specific subjects and may select their students through aptitude tests (e.g., Varjo & Kalalahti, 2019)—have been proposed to act as implicit tracks within the Finnish comprehensive school system (Berisha & Seppänen, 2017). Importantly, these classes must follow the same national core curriculum as general classes without an emphasis, and therefore, the Finnish context provides a unique setting for studying the effects of tracking as the instructional variation across tracks is less pronounced than in many other countries that utilize tracking.

1.1. Teachers' role in shaping the development of children's motivational beliefs during elementary school

Children's positive motivational beliefs are known to decrease over time (for review, see Muenks et al., 2018). Suggested explanations for the detected decline include cognitive developmental changes (e.g., Demetriou et al., 2011) as well as age-related changes in children's typical environments (e.g., Butler, 2005).

A prevailing explanation attributes children's declining motivation to Piaget's cognitive development theory, suggesting that young children are initially overly optimistic about their abilities due to an immature and undifferentiated understanding of competence (Piaget & Inhelder, 1956; Skinner et al., 1998). As children mature, their understanding of concepts such as ability and effort becomes more refined (Chapman & Skinner, 1989; Little et al., 2002; Nicholls, 1978). However, this canonical explanation has been challenged by evidence showing environmental impacts on children's self-appraisals (Butler, 2005; Cimpian, 2017). Research indicates that in familiar contexts, young children's competence judgments may seem more advanced than expected, influenced by age-related changes in children's typical experiences and contexts rather than solely by cognitive development (Butler, 2005; Eccles et al., 1993).

Self-appraisals develop in interaction with one's environment (Urdan & Schoenfelder, 2006). For the formation of achievement motivation, achievement situations are especially important, as self-appraisals of competence develop reciprocally with experiences of achievement (Pajares & Schunk, 2001; Skinner et al., 1998). When children start formal schooling, they are for the first time exposed to an environment in which they are required to acquire and demonstrate specific skills or understandings, and their involvements and achievements are constantly evaluated (Butler, 2005). Experiences of success or failure in various achievement situations, feedback from teachers, and comparisons with classmates shape individuals' understanding of their own competencies, making school an important environment influencing children's motivational beliefs (Skaalvik & Skaalvik, 2002; Urdan & Schoenfelder, 2006).

Teachers are an important source of competence-related feedback for children at school. As Skaalvik and Hagtvet (1990, p. 295) have concluded, "*Teachers are significant others for most students, thereby providing important clues about academic performance. The reflected appraisals from the teachers may have a strong impact on the students' self-concept of ability.*" Prior studies have shown that teacher's beliefs or perceptions of students' abilities and motivation, along with their expectations, shape students' self-perceptions as learners (e.g., Bohlmann & Weinstein, 2013; Pesu et al., 2016). Teachers' expectations are sometimes proposed to act as a self-fulfilling prophecy, becoming confirmed in children's own behavior and attitudes (Jussim et al., 2009; Weinstein, 2002) and this same mechanism has been suggested for teachers' ability perceptions too (Upadyaya & Eccles, 2015). For example, in studies by Upadyaya & Eccles, utilizing longitudinal data from 849 primary school children and their teachers, the authors noted that teachers' beliefs about students' effort and potential performance predicted children's interest in math over the school years (Upadyaya & Eccles, 2014) as well as children's self-concept in math (Upadyaya & Eccles, 2015). The influence of teachers on children's self-perceptions has been shown to be especially important during the first years of schooling (Pesu et al., 2016), with evidence also suggesting an escalating impact throughout the elementary-school years (Spinath & Spinath, 2005).

Teachers convey their expectations and beliefs to students in multiple conscious and unconscious ways including explicit evaluations (e.g., school marks, assessment on school assignments) and feedback in achievement situations. In addition, teachers' emotional reactions such as praise, pity, or blame (Georgiou et al., 2002; Graham, 1984) and the methods they employ in the classroom (e.g., Bohlmann & Weinstein, 2013; Marshall & Weinstein, 1984; Weinstein, 2002) are important in molding children's understanding of themselves. Craven et al. (1991) have suggested that the feedback a teacher gives to a child influences their later self-perceptions through a process in which the child internalizes the teacher's feedback in a specific task or situation as a feeling of success ("I did well in this"). These experiences are later generalized into a feeling of competence in that same domain ("I am good in mathematics").

Studies examining the impact of teachers have typically defined expectations as the judgments teachers make about students' future achievement and behavior, based on their current knowledge of the students (Riley & Ungerleider, 2012; Timmermans et al., 2016). In comparison, teachers' beliefs[2] or perceptions of students refer more to the present-moment perceptions that teachers have about children (Gentrup & Rjosk, 2018; Upaduay & Eccles, 2014). Thus, the key difference between these concepts lies in their temporal dimension, with expectations oriented towards the future and beliefs or perceptions grounded in the present moment. Research has suggested that teacher's expectations may be shaped by their perceptions of students' behavior at school. For example in their study of 5316 students from 469 classes in Dutch primary education, Timmermans et al. (2016) investigated the relationship between teachers' perceptions of student attributes (i.e. working habits, popularity, self-confidence, student–teacher relationships, and classroom behavior) and teachers' expectations and showed that teachers had higher expectations for students whom they perceived as self-confident and hard working.

Teachers' expectations and perceptions of children have shown to be influenced by a range of socio-demographic factors, including the child's social background and gender (e.g., Butler & Hasenfratz, 2017; Gunderson et al., 2012; Mullola et al., 2012; Riegle-Crumb & Humphries, 2012; Timmermans et al., 2016; Wang et al., 2018). Studies have shown that teachers' gendered perceptions of ability may contribute to the consistent effects of gender on children's self-beliefs (Herbert & Stipek, 2005). Overall, girls tend to exhibit a more positive academic motivation pattern, marked by strong effort and high achievements, but often accompanied by lower self-confidence (Butler & Hasenfratz, 2017; Parker et al., 2018). Interestingly, despite extensive research demonstrating that girls surpass boys in various academic subjects (Voyer & Voyer, 2014), boys consistently assess their own abilities as higher than girls do (Parker et al., 2018). This phenomenon is likely influenced by societal expectations and gender stereotypes propagated by significant

[2] The term 'teacher beliefs' also encompasses teachers' personal motivational beliefs, such as their views on the nature of ability (fixed versus malleable), which have been shown to influence their teaching practices and subsequently affect students' self-beliefs (e.g., Stipek et al., 2001).

adults, such as teachers and parents (Butler & Hasenfratz, 2017; Eccles, 2011). Furthermore, stereotypical expectations contribute to domain-specific gender differences, with boys generally exhibiting stronger self-concepts in traditionally masculine subjects like mathematics and physical sciences, while girls excel in verbal subjects (Marsh, 1989). In Finland, gender differences in education have been especially pronounced and favoring girls (Leino et al., 2019). For instance, in the latest PISA 2022 assessment, Finnish girls outscored boys in both reading and even in mathematics (OECD, 2023a), a subject where boys typically outperform girls in many countries (Early et al., 2020). Finnish girls also outperformed boys in creative thinking in the same assessment (OECD, 2023b).

A central finding from previous research, relevant to this study, is that teachers' expectations and beliefs about their students are linked to student grouping practices (e.g., Bohlmann & Weinstein, 2013; Kuklinski & Weinstein, 2001; Marshall & Weinstein, 1984). For instance, in a mixed-method study by Mazenod et al. (2019), which combined quantitative survey data from 597 teachers across 82 schools with qualitative in-depth interviews from 34 teachers, it was discovered that teachers perceived low-track students through a nurturing discourse. This discourse became explicit in teachers' lowered expectations regarding learning (i.e., independence, pace, and depth of learning) for students in lower attainment groups. Van Houtte et al. (2013) examined the effect of tracking on secondary school teachers' perceptions of students' cognitive capacity, effort, and diligence in doing homework. Analyzing data from 6545 students across 46 secondary schools in Flanders, they found that while teachers' perceptions of students' capabilities and diligence in homework varied by tracks, the differences between tracks were primarily explained by students' background features, indicating a selection effect. Contrary to this, Kelly and Carbonaro (2012), using a nationwide representative longitudinal sample of 8th graders from the United States and focusing on students with discrepant track placement, uncovered an independent tracking effect on teachers' expectations. They concluded that track placements affected teacher expectations above and beyond student achievement and other characteristics. Furthermore, also in Van Houtte's et al. (2013) study, one independent effect of tracking regarding effort, was observed, which could not be explained by selection. Interestingly, contrary to intuitive expectations, this effect favored low-track students, indicating that teachers perceived low-track students as investing more effort in their schoolwork compared to students in top tracks.

1.2. Influence of tracking on learning outcomes and motivation

Effects of tracking, streaming, or ability-grouping have been widely studied in the educational sciences, ranging from the economics of education to educational sociology and psychology (e.g., Chiu et al., 2017; Collins & Gan, 2013; Duflo et al., 2011; Kindermann, 2016). Practices of tracking differ widely between and within nations and school systems (Dupriez et al., 2008; OECD, 2016), making it difficult to extrapolate the findings from one context to another and draw general conclusions of the impact of tracking. In addition, examining the effects of tracking can be methodologically challenging, due to many confounding factors and simultaneous effects (e.g., Televantou et al., 2021).

Effects of tracking can be scrutinized from several perspectives. At the individual level, the way students are assigned to schools or teaching groups within schools plays a pivotal role in shaping their learning environment, which may affect learning in multiple direct and indirect ways (see Collins & Gan, 2013). At the societal level, tracking practices may have significant implications for educational equality (e.g., Hanushek & Wössman, 2006; Wößmann, 2009). The practice of educating all children together in a comprehensive school system versus dividing them into separate tracks stands as a critical choice that may either advance or impede the pursuit of educational equality.

Pallas et al. (1994) distinguished instructional (i.e., quantity, quality, and differing instructions in different tracks), social (i.e., social setting, class composition), and institutional (i.e., how different tracks are valued and seen) mechanisms of tracking. By creating more homogeneous student groups, tracking may have a direct impact on learning, as, at the instructional level, tracking allows teachers to customize their teaching strategies to address the distinct needs of each group. This customization of instructions and curricula is suggested to be a crucial element in explaining the achievement differences between tracks (Gamoran, 2010). In addition, tracking may also have indirect effects on students learning through the distinct social environments and class compositions it creates. Such indirect pathways can include peer influences (e.g., Cooc & Kim, 2017; Kindermann, 2016; Wentzel & Muenks, 2016; Zimmer, 2003) or the distinct expectations, discussed in the previous section, that students' significant others (i.e. teachers and parents) set for students studying on different tracks (e.g., Andersen, 2018). Our paper centers on the latter aspect, exploring how tracking shapes teachers' view of students' academic schoolwork skills and how these perceptions, in turn, predict students' motivational self-beliefs and test performance.

Empirical studies on tracking effects vary by outcomes examined and focus, ranging from individual to school-level impacts on learning, achievement, educational equality, and affective outcomes (see Gamoran, 2010; Hattie, 2009 for achievement; Gamoran, 2010; Terrin & Triventi, 2023 for educational equality; Belfi et al., 2012 for affective outcomes). Most studies on tracking have primarily focused on academic achievement as an outcome (e.g., Chiu et al., 2017; Collins & Gan, 2013; Steenberg-Hu et al., 2016). A tentative summary of the existing research indicates that at the school system level, the impact of tracking on student achievement tends to be minimal or even statistically non-significant (Hattie, 2009). On the other hand, there is a body of evidence indicating that early tracking may contribute to greater educational inequality (e.g., Dupriez et al., 2008; Hanushek & Wössman, 2006). For example, in their comprehensive meta-analysis drawing from 53 publications on the effects of tracking on students' achievement Terrin and Triventi (2022) conclude that overall tracking does not seem to enhance the efficiency of the educational system (i.e. increase achievement at the school level). Conversely, it appears to widen the dispersion of achievement outcomes and amplify the influence of family background on student achievement, thereby increasing educational inequality.

However, alongside these findings, a substantial number of studies have explored tracking's effects at the individual student level and shown that tracking can indeed have significant effects on students' abilities or achievement (e.g., Becker et al., 2012; Retelsdorf et al., 2012). Some studies have shown tracking to be beneficial for all students (e.g., Collins & Gan, 2013; Duflo et al., 2011; Tieso, 2003), whereas others have suggested that the effects of tracking may differ by students' ability level (Gottfried, 2014) or other individual attributes, such as gender or ethnic background (Hoxby, 2000) as well as the type of tracking (e.g., Chmielewski, 2014; Steenbergen-Hu, 2016).

Aside from achievement, academic self-concept has been the main outcome studied in relation to tracking effects (e.g., Marsh & Seaton, 2015; Trautwein et al., 2006). Research has shown that due to social comparison processes, studying in academically demanding high-track environment in which the average achievement level is high, may have detrimental effects on children's academic self-concept (i.e. Big-Fish-Little-Pond-Effect, BFLPE) (e.g., Fang et al., 2018; Marsh & Seaton, 2015). However, although with less robust evidence, studies have suggested that negative effect of BFLPE might be counterbalanced by comparisons with students in other tracks (i.e. the Reflected Glory Effect, RGE), which may enhance the self-concept of students in higher tracks (Marsh et al., 2000; Trautwein et al., 2005). The presence of these simultaneous effects can explain the mixed findings, suggesting that low-achievers' self-concept may benefit from tracking in some cases (Dupriez et al., 2008), while in others, it may suffer from being placed in a low-track environment (Francis et al., 2017). Chmielewski et al. (2013) have suggested that the type of tracking could elucidate these

divergent findings. Using PISA data from 99 000 15-year-olds across 3400 schools in 20 countries, they discovered that the type of tracking influenced the relationships between different effects, leading to variations in students' self-concepts depending on the type of track they were in. In course-by-course tracking, where students are grouped only for specific courses, the RGE outweighed the negative effect of the BFLPE, resulting in higher self-concepts for students in higher tracks and lower self-concepts for those in lower tracks. However, in between-school and within-school tracking, the negative BFLPE was more prominent. The authors attribute this difference to variations in reference groups used for comparison. In course-by-course tracking, the reference group remains the entire age cohort in the school, leading student to compare themselves against the overall achievement level of their respective tracks (high or low). Consequently, students in high-track courses may bolster their self-concept based on their track placement, whereas those in low-track courses may experience diminished self-perceptions. Conversely, in between-school and within-school tracking, the reference group consists solely of peers within the same track, leading students in higher tracks to suffer from high-achieving peer comparisons. However, for low-track students, the consequences of tracking are less harmful than in course-by-course tracking.

In addition to academic self-concept, tracking effects have been studied in relation to other motivational outcomes, but the field of research is much more scattered. For example, Carbonaro (2005) studied the effects of curricular tracking on lower secondary school students' effort and achievement and found students in higher-achieving tracks to exert more effort than students in lower-achieving tracks. Then again, Butler (2008) showed that tracking increased ego-orientation (i.e., importance of showing off in front of others) and work-avoidance goals (i.e., avoiding effort exertion) among elementary-school children, indicating that tracking based on ability may increase the salience of social comparison processes in the classroom.

1.3. Implicit tracking in Finland: classes with special emphasis

Much of the prior research on tracking has been conducted in Anglophone countries, and therefore their context differs notably from this study. Instructional variation across tracks has been suggested as one of the main reasons explaining achievement gaps between tracks (Gamoran, 2010). In this paper, we focus on examining the effects of tracking in the Finnish context, which offers an interesting case for a tracking study and enables us to study the effects of tracking in a context in which the instructional variation between tracks is minor.

The Finnish basic education system has typically been presented as a model example of a non-tracked system, in which all children study according to the same national core curriculum[3] and get the same educational opportunities for upper-secondary education (Seppänen, 2003; Wößmann, 2009). Less attention, at least internationally, has been given to classes with a special emphasis, which were established in the 1990s and which have been argued to act as an implicit tracking system within the Finnish comprehensive school system (Berisha & Seppänen, 2017; Seppänen ym., 2023). Since their introduction, classes with a special emphasis have gained popularity, particularly in large cities, providing a variety of options and choices. These options span a range of academic subjects, such as mathematics, science, and languages, to arts including music, visual arts, and sports. Research indicates that in some cities, about 30% of students are enrolled in classes with a special emphasis (Seppänen et al., 2015). Typically, students have the opportunity to enroll in these specialized classes at the start of comprehensive school at age seven, or before entering the 3rd or 7th grades.

Classes with a special emphasis provide additional teaching (i.e., 1–2 h/week) in a specific subject (e.g., music, mathematics, sports, a foreign language), and have the right to select students with aptitude tests (Varjo & Kalalahti, 2019). Prior research has indicated that selection into classes with a special emphasis is related to social class, and while academic achievement is not a criterion for selection, high-achieving children from highly-educated families are typically overrepresented in these classes (Kosunen & Seppänen, 2015). Therefore, researchers have argued that classes with a special emphasis share many similarities with ability-grouping or tracking (Berisha & Seppänen, 2017). As selection into classes with a special emphasis is not based on prior achievement, but on other aptitudes (Varjo & Kalalahti, 2019), we have suggested elsewhere (Koivuhovi, 2021) that the term opt-in tracking within schools (see Trautwein et al., 2005) might best describe the Finnish system. Selection criteria vary according to the emphasized subject and between municipalities (Seppänen, 2003). Usually, schools arrange aptitude tests where children can show their talent and interest in the subject.

Even though classes with a special emphasis have been researched in Finland since their establishment, few studies have analyzed their developmental effects on students. In our prior studies (Koivuhovi, 2021; Koivuhovi et al., 2022) investigating the development of children's motivation in classes with and without a special emphasis, we found no significant effects of tracking. However, regarding the development of achievement, prior research is even more limited. Classes with a special emphasis have been shown to increase within-school differences, as high-achieving children are grouped together (e.g., Berisha & Seppänen, 2017; Kosunen et al., 2020), but it is unclear whether studying in a class with a special emphasis actually influences the development of learning, or whether the noted differences between classes are only due to selection effects.

2. Present study

The aim of the present study is to examine the influence of teachers' perceptions and tracking on changes in children's motivational self-beliefs and achievement. Our study is organized around three objectives. Firstly, we investigate the role of teachers' perceptions of children's academic schoolwork skills in the change of children's motivational self-beliefs and achievement. Secondly, we explore how gender, mother's education level,[4] and class type (class with a special emphasis versus general class without an emphasis) relate to teachers' perceptions of children's academic schoolwork skills. Regarding general relations between teachers' perceptions and children's self-beliefs and achievement, as discussed above (e.g., Pesu et al., 2016; Spinath & Spinath, 2005; Upaduay & Eccles, 2014), we expect (Hypothesis 1) teachers' perceptions of children's academic schoolwork skills to be significant positive predictors of change in children's later self-beliefs. In addition, based on prior research (e.g., Andersen, 2018; Timmermans et al., 2016; Wang et al., 2018), we expect that teachers' perceptions of students will differ by students' gender, mothers' education levels and class types (Hypothesis 2).

Our third aim focuses on the role of tracking in the change of children's motivational beliefs and achievement. Building upon our previous investigations (Koivuhovi et al., 2022), we expect (Hypothesis 3) that class type will not be a significant predictor of changes in children's motivational beliefs once covariates (gender and maternal education) are accounted for. Similarly, we expect not to find effects on achievement as differences are often attributed to variations in curricula between tracks (Gamoran, 2010), and the Finnish education system mandates a uniform curriculum across all tracks (Seppänen, 2003). However, considering prior international research on the effects of

[3] Although schools are also allowed to formulate their own local curricula, the obligations set by the national core curriculum leave them little space for any major changes (Kantasalmi & Kupiainen, 2021).

[4] Mother's education level has been used in the analyses as a proxy for socio-economic status, as it has been shown to be an effective indicator of socio-economic status (Jalovaara & Andersson, 2018).

tracking (e.g., Becker et al., 2012; Carbonaro, 2005; Chiu et al., 2017; Retelsdorf et al., 2012), it would be conceivable to propose an alternative hypothesis suggesting that tracking could significantly influence children's achievement and motivation. Therefore, our third hypothesis is formulated with cautious optimism, acknowledging the potential for other than expected outcomes.

Our analyses proceed in stages. Initially, we examine the overall relationships between teachers' perceptions and children's self-beliefs. Subsequently, we incorporate achievement into the model to assess its impact on these relationships. Finally, we introduce covariates, including class type, gender, and mother's education, to investigate their effects on both teachers' perceptions and children's self-beliefs and achievement.

3. Methods

3.1. Data collection procedure and participants

The data were drawn from a nine-year longitudinal study in which children's cognitive abilities and motivational beliefs were followed from the beginning to the end of comprehensive school. In the beginning of the study, 16 schools in a large municipality in Southern Finland were randomly selected. The sample consisted of all first graders ($n = 744$) in the selected schools. The sample size increased in the beginning of the fourth and seventh grades, as some children changed classes or schools, and the follow-up was extended to cover also the new classmates of our participants (for detailed description of the data, see Koivuhovi, 2021). For this paper, we utilized data from grades four to six, during which children were taught by the same teacher, a condition essential for the purposes of our study.

The participants ($n = 1038$) came from 55 classes but small classes specifically designed for students with special education needs were not included in the analyses.[5] Therefore, participants of this study came from 47 classes, of which 11 with a special emphasis ($n = 291$). Classes with a special emphasis included emphases on languages, arts, music, and physical education. Due to this variety and the limited number of classes, it was not feasible to examine different classes separately. Therefore, the analyses were conducted based on differences in class types (i.e. class with or without a special emphasis). Children who studied in classes with a special emphasis had started this special track from the beginning of third grade. Teachers had thus taught students for at least a year when assessing their academic schoolwork skills at the beginning of the fourth grade. The teaching groups were arranged so that the same teacher taught the children in most subjects.

Thanks to the diligent efforts of the research coordinator overseeing data collection, the occurrence of missing data between waves was minimal. Sample sizes varied slightly across different years and variables. For instance, at grade four, achievement test scores were available for 897 students, and teacher evaluations of child's effort were available for 911 students. At grade six, achievement test scores were available for 912 students, while teacher evaluations of child's effort were recorded for 865 students. For both years combined, data included achievement scores for 833 students and teacher evaluations for 774 students. At the item level, the amount of missing data was moderate, ranging from zero to 17 % per item. The average amount of missingness per item was 11.86 % which was taken into account in the analyses using Mplus's default technique for handling missing values, namely Full-Information Maximum Likelihood (FIML).

Mean age of the participants at grade four was 9.6 years, and 52 % of them were girls. Children who studied in classes with a special emphasis had mothers with higher education levels compared to those in general education classes ($\chi 2$ (3, $N = 940$) = 70.420, $p < .001$). Additionally, at the fourth grade level, significant differences were observed in test proficiency between the class types: students in classes without an emphasis had a mean score of 17.5 ($SD = 6.30$), while those in classes with a special emphasis scored higher, with a mean of 18.5 ($SD = 6.40$; $F = 4.234$, $p = .039$).

Throughout the whole study, ethical standards described by the Finnish Advisory Board on Research Integrity (see, TENK, 2012) were met. Educational authorities of the municipality reviewed the proposal of the study and gave permission for data collection. Families were informed about the study and participation was voluntary. Although families had the option to opt out, very few chose to do so.

3.2. Measures

Teachers' perceptions of children's academic schoolwork skills were collected from teachers and measured with four items developed specifically for this study. The items included: "Child focuses on teaching during lessons"; "Child concentrates on their schoolwork"; "Child takes care of their own things at school"; "Child tries hard at school"). Items were answered on a seven-point Likert scale (1 = not true at all – 7 = very true).

Children's motivational beliefs were assessed with scales based on Skinner et al.'s (1988) action control theory, in which action is conceptualized as a threefold construct, comprising the agent, aims, and means of action. An individual (i.e., the agent) possesses different kinds of beliefs (i.e., means-ends beliefs, agency beliefs, control expectancy) concerning each part of the action, and these beliefs shape their actions and strivings (Skinner et al., 1988, 1990). In the present study, we focused on children's agency beliefs and control expectancy beliefs, which are central in describing motivated behavior. Agency beliefs are individual's self-beliefs as an agent in relation to the possible means of action (Skinner et al., 1988, 1990). Thus, agency beliefs pertain to students' self-perceptions as learners (Little et al., 2001), while control expectancy beliefs relate to individuals' expectations regarding their ability to achieve desired goals (Skinner et al., 1988, 1998). Within the school context, these beliefs encompass expectations about, for example, the likelihood of academic success or obtaining good grades.

For this study, the scales were specifically modified and translated to fit the Finnish context. Agency beliefs were measured using two scales: one assessing beliefs about ability (e.g., "I am a clever and able student") and another assessing beliefs about effort (e.g., "I work hard at school"). Control expectancy beliefs regarding academic success were measured with a single scale (e.g., "I can get good marks at school, if I want to"). Each scale consisted of three items, and responses were answered on a seven-point Likert scale ranging from 1 (not true at all) to 7 (very true).

The Achievement test score was a composite score derived from various cognitive tasks measuring reading comprehension skills, mathematical thinking skills and general reasoning skills. All together the composite score was calculated based on 40 items and the internal consistency of the measure was good (for grade 4 $\alpha = .77$ for grade 6 $\alpha = .84$). Each item was coded dichotomously as correct or incorrect, with the composite score reflecting the total number of correct responses.

Reading comprehension skills were assessed with two tasks. The first, based on the framework by Kintsch and van Dijk (1978), involved evaluating 16 statements related to a one-page text. The second task required interpreting a short text related to everyday life, assessing children's ability to understand, analyze, and interpret written information through four multiple-response items (Hautamäki et al., 2002).

Mathematical thinking skills were assessed with two tasks: First test was a modified version of Mental Arithmetic test by Wechsler (1981) which involved solving verbal math problems within a limited time. For

[5] These students receive the most intensive level of support, known as 'special support,' under Finland's three-tier support system (Hienonen, 2020). As such, they follow individualized education plans, and therefore their education may not necessarily be comparable with that of other students. Including these students in our study would have introduced substantial variability and would not have been representative of the broader student population. Therefore, we decided to exclude these students from our analysis.

example, students responded to problems such as, "The merchant had 25 cans of milk and he sold 14 of them. How many cans did he have left?" read aloud by the teacher. The second test was the Hidden Arithmetic Operators test by Demetriou et al. (1991, 1996), where students identified hidden operators necessary to solve equations with each item containing one to four hidden operators (e.g., [(3 a 2 b 4) c 3 = 7, where a/b stands for addition (+), subtraction (−), multiplication (•), or division (÷)]).

General reasoning skills were assessed through two tasks: the Piagetian water-level task (Piaget & Inhelder, 1956; Hautamäki, 1985), which required students to indicate water levels in tilted bottles, and Bond's Logical Operations Test (Bond, 1995), which involved choosing correct answers from multiple options.

The independent variables in the analyses included gender (boy or girl), class type (class with or without a special emphasis), and maternal education level (collected from parents; 1 = compulsory education; 2 = upper secondary education; 3 = first-cycle tertiary education i.e. Bachelor's degree or similar; 4 = second-cycle tertiary education and above i.e. Master's degree or similar).

3.3. Data analysis

Data were analyzed with Mplus version 8.6 (Muthén & Muthén, 2018). Before the actual analyses, measurement invariance of the latent factors over time was examined with separate models for each construct (see Table 1). Invariance testing was conducted by imposing restrictions to the model parameters and then comparing fit indices of different models (Widaman et al., 2010). This process began with configural invariance, where no constraints are applied, advanced to metric invariance by equalizing factor loadings, and concluded with scalar invariance through the equalization of item intercepts, following the methodology outlined by Millsapp and Cham (2013). The criteria for evaluating model fit and determining the significance of any decline in fit were informed by Chen's (2007) recommendations. Accordingly, for metric invariance, an acceptable deviation was defined as a change in CFI of less than 0.010, accompanied by a change in RMSEA of less than 0.015 or a change in SRMR of less than 0.030. For scalar invariance, acceptable criteria were a change in CFI of less than 0.010, with a change in RMSEA of less than 0.015 or in SRMR of less than 0.010 (Chen, 2007). As indicated in Table 1, the analyses confirmed an adequate level of measurement invariance over time in each pair of latent factors.

Table 1
Measurement invariance of the latent factors over time.

Tested model	*χ2*	*df*	*CFI*	*RMSEA*	*SRMR*
Teachers' perceptions					
Configural invariance	40.045	15	0.996	0.041	0.017
Metric invariance	52.501	18	0.994	0.044	0.032
Scalar invariance	81.545	21	0.990	0.054	0.042
Partial scalar invariance (intercepts of item 4 free)	53.863	20	0.994	0.041	0.033
Agency beliefs of ability	*χ2*	*df*	*CFI*	*RMSEA*	*SRMR*
Configural invariance	19.963	8	0.995	0.039	0.020
Metric invariance	23.078	10	0.995	0.037	0.030
Scalar invariance	135.636	12	0.949	0.103	0.071
Partial scalar invariance (intercepts of item 2 free)	24.489	11	0.994	0.036	0.030
Agency beliefs of effort	*χ2*	*df*	*CFI*	*RMSEA*	*SRMR*
Configural invariance	5.753	8	1.00	0.000	0.011
Metric invariance	7.192	10	1.00	0.000	0.013
Scalar invariance	25.301	12	0.995	0.034	0.030
Partial scalar invariance (intercepts of item 1 free)	7.315	11	1.00	0.000	0.013
Control expectancy beliefs of success	*χ2*	*df*	*CFI*	*RMSEA*	*SRMR*
Configural invariance	19.880	8	0.994	0.039	0.021
Metric invariance	22.510	10	0.994	0.036	0.028
Scalar invariance	24.898	12	0.994	0.033	0.031

The research hypotheses were assessed with cross-lagged panel models, including measures from two time points (grades four and six). With these cross-lagged panel-models we regressed a set of variables measured in grade six on a set of predictors measured in grade four, while controlling for the stability of the all measures. Thus, we estimated to which degrees our grade-four-measures predict a *change* in the rank order of the grade-six-measures. Analyses proceeded in stages, so that a model with latent factors in two time points (i.e., teachers' perceptions and action-control beliefs) was first analyzed, to investigate the relations between teachers' perceptions and children's motivational self-beliefs. Following this, first achievement test scores from two time points, and finally covariates, were included in the model.

The datapoints of our study are nested in classrooms. Since datapoints within a classroom are on average more similar to each other than datapoints between different classrooms (the Intra-Class-Correlations of our variables of interest ranged from 0.005 for control expectancy beliefs to 0.185 for achievement), we had to account for this nested structure in our statistical analyses to estimate standard errors correctly. We used the type = complex feature of Mplus to account for this nested data structure. Our models converged successfully but Mplus issued the warning that "Standard errors may not be trustworthy". This warning was most likely due to the model having more parameters than the number of clusters available for analysis. To verify the accuracy of the standard error estimations, we conducted likelihood ratio tests on each path coefficient in our three separate models, following the approach recommended by Gonzalez and Griffin (2001). We compared a reduced model—where a specific path coefficient, A, was fixed at zero—to the full model. If the likelihood ratio test showed no significant difference in fit between the models, and the p-values of the tested parameter matched those in the full model, it indicated correct standard error estimations. We based our substantive conclusions solely on the p-values from these likelihood ratio tests, which closely aligned with those from the full models.

4. Results

All the analyzed models fit the data well (Model fit for the final model with covariates: $\chi2 = 612.174$; $df = 372$; $p <. 00$; CFI = 0.976; TLI = 0.970; RMSEA = 0.027; SRMR = 0.033).[6] As we expected (H1), teachers' perceptions of children's academic schoolwork skills significantly predicted children's later self-beliefs, whereas children's motivational beliefs were not significant predictors of teachers' perceptions (see Fig. 1, Table 2 all models). This indicates that teachers' perceptions of children's academic schoolwork skills shape children's developing self-perceptions, but not vice versa. Adding achievement and covariates to the model did not change these relations and analyses of the full model showed that teachers' perceptions of children's academic schoolwork skills predicted children's later self-beliefs, even after actual achievement and other background variables (gender, mothers education, class type) were controlled (see Table 2, Model3).

Relationships between children's achievement test scores and teachers' perceptions gave evidence of a reciprocal model (see, Guay et al., 2003). In other words, students' achievement test scores at grade four significantly predicted teachers' perceptions at grade six, and vice versa (see, Fig. 1, Table 2, Model 2). However, the introduction of covariates altered these relationships (see Table 2, Model 3). Full model with covariates, indicate that teacher's perceptions of students have an

[6] In order to obtain a sufficient model fit, residual correlations between same items in different time points were allowed for teacher perceptions. Model fit for the first model without achievement and covariates was ($\chi2 = 429.134$; df = 282; $p <. 000$; CFI = 0.984; TLI = 0.981; RMSEA = 0.024; SRMR = 0.032) and ($\chi2 = 481.562$; df = 318; $p <. 000$; CFI = 0.983; TLI = 0.980; RMSEA = 0.023; SRMR = 0.032) for the second model with achievement.

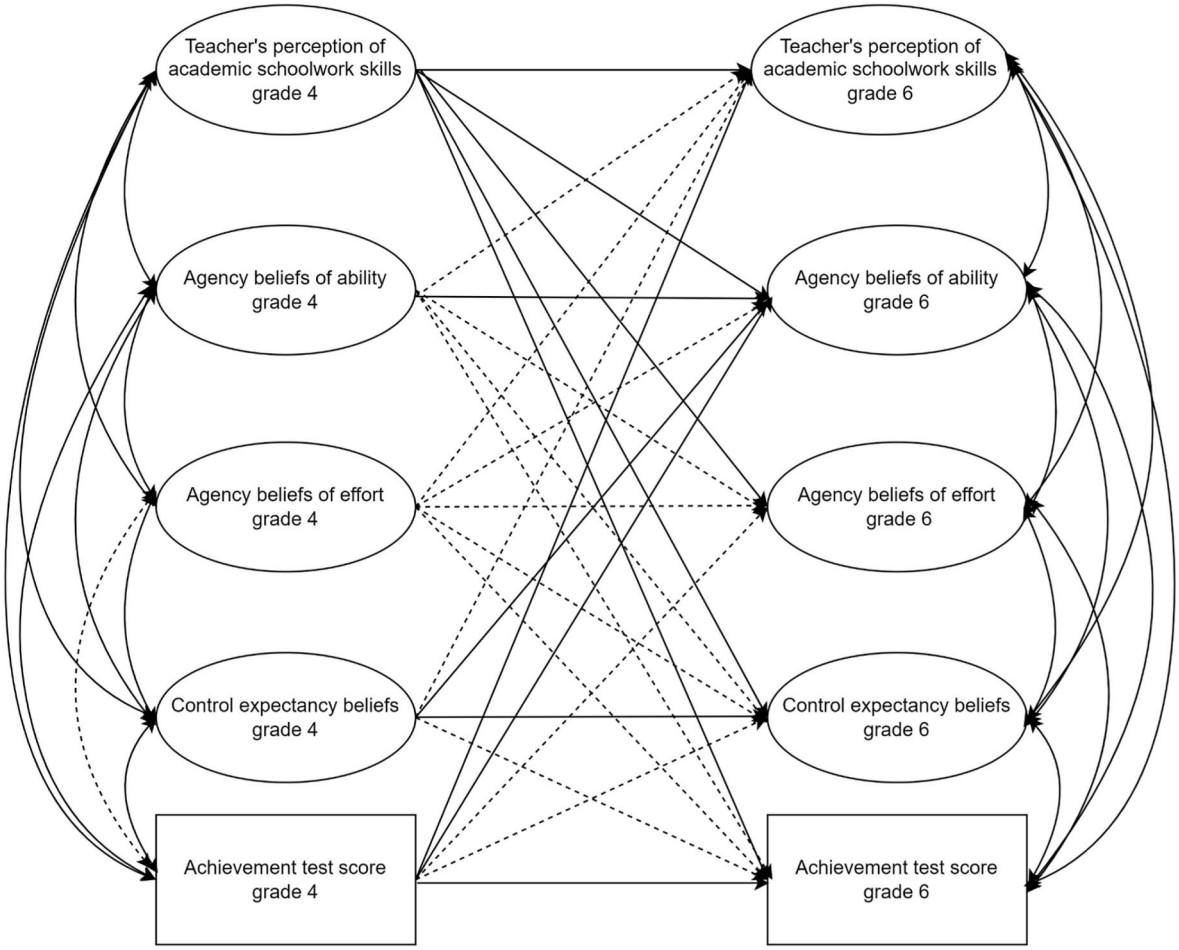


Fig. 1. *Cross-lagged panel model with motivational factors and achievement test scores (Model 2). Note.* Statistically significant paths are visualized with solid, non-significant paths with dashed lines.

effect on the change of their achievement but achievement itself is not influencial in how teachers perceive students academic schoolwork skills. Instead, gender, mother's education, and class type significantly influenced how teachers perceived children. Regarding achievement, precisely mother's education level appeared to account for the diminishing influence of achievement on teachers' perceptions when covariates were taken into consideration. Mother's education was strongly connected to child's achievement test scores at both time points (Effect of mother's education on achievement at 4th grade: $\beta = 0.33$, $SD = 0.05$, $\Delta\chi 2 = 75.702$, $p < .000$; at sixth grade $\beta = 0.19$, $SD = 0.030$, $\Delta\chi 2 = 35.025$, $p < .000$) and therefore adding it to the model wiped away the effect of achievement on teacher's perceptions.

Girls were consistently perceived more positively by teachers than boys, as indicated by significant effects at both fourth grade ($\beta = 0.56$, $SD = 0.09$, $\Delta\chi 2 = 51.623$, $p < .000$) and sixth grade ($\beta = 0.29$, $SD = 0.08$, $\Delta\chi 2 = 15.773$, $p < .000$). Similarly, children with highly educated mothers were also viewed more positively, shown by the effects at fourth grade ($\beta = 0.19$, $SD = 0.04$, $\Delta\chi 2 = 22.447$, $p < .000$) and at sixth grade ($\beta = 0.07$, $SD = 0.04$, $\Delta\chi 2 = 3.914$, $p = .048$). This result indicates, as we expected (H2), that gender and mother's education[7] were significant factors in explaining not only the positive teachers' perceptions in the beginning of the follow-up period but also in explaining the change in the perceptions.

Class type was a significant predictor of teachers' perceptions, but interestingly, its effect on evaluations was different at different time points. At fourth grade, studying in a general class predicted more positive teacher perceptions, whereas at sixth grade, conversely, studying in a class with a special emphasis predicted higher teacher perceptions (Effect of class type on teacher's perceptions at 4th grade: $\beta = -0.31$, $SD = 0.14$, $\Delta\chi 2 = 10.536$ $p < .000$; at sixth grade $\beta = 0.28$, $SD = 0.14$, $\Delta\chi 2 = 11.369$, $p < .000$.

Contrary to our expectations (H3), class type was associated with

[7] It must be noted that in the type complex model, the effect of mother's education on teacher's perceptions was the only one that differed from chi-square model comparison results. In the type complex model, the effect of mother's education on teacher's perceptions did not reach statistical significance ($\beta = 0.07$, SD = 0.04, p = .076) whereas in the chi-square difference test between the non-restricted model (all paths evaluated) and the restricted model (teacher's perceptions at 6th grade regressed on mother's education set as zero), the difference between models was significant ($\Delta\chi 2 = 3.914$, p = .048). Given these nuances, this specific result should be approached with caution and considered as indicative rather than strongly conclusive.

Table 2
Standardized model estimates and chi-square model comparisons of different models.

Model	Dependent variables (T2)																			
Variable T1	TP (Teacher perceptions)				AA (Agency belief of ability)				Agency: Effort				Control expectancy				Achievement score			
Model 1	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*
TP	**0.60**	**0.06**	**58.716**	**0.000**	**0.23**	**0.06**	**20.168**	**0.000**	**0.20**	**0.06**	**13.882**	**0.000**	**0.15**	**0.06**	**8.362**	**0.003**	–	–	–	–
AA	0.06	0.09	0.712	0.398	**0.41**	**0.15**	**9.622**	**0.001**	0.11	0.11	1.059	0.303	0.20	0.14	2.292	0.130	–	–	–	–
AE	0.02	0.10	0.46	0.497	−0.26	0.15	3.782	0.051	0.14	0.12	1.286	0.256	−0.12	0.12	1.121	0.289	–	–	–	–
CE	−0.02	0.04	0.534	0.464	0.15	0.06	7.129	0.007	0.01	0.05	0.381	0.537	**0.25**	**0.06**	**18.665**	**0.000**	–	–	–	–
Model 2	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*
TP	**0.55**	**0.07**	**41.513**	**0.000**	**0.20**	**0.06**	**11.573**	**0.000**	**0.22**	**0.06**	**13.456**	**0.000**	**0.14**	**0.06**	**5.664**	**0.017**	**0.20**	**0.05**	**20.042**	**0.000**
AA	−0.01	0.11	0.407	0.523	**0.38**	**0.15**	**6.878**	**0.008**	0.15	0.12	1.509	0.219	0.19	0.14	1.846	0.174	0.08	0.10	0.618	0.431
AE	0.11	0.12	1.151	0.283	−0.22	0.16	2.237	0.134	0.09	0.13	0.754	0.385	−0.11	0.13	0.935	0.333	−0.13	0.11	1.187	0.275
CE	−0.03	0.04	0.727	0.393	0.14	0.06	6.639	0.009	0.02	0.05	0.444	0.501	**0.25**	**0.06**	**18.804**	**0.000**	0.08	0.05	3.521	0.060
ACH	**0.12**	**0.05**	**6.417**	**0.011**	0.08	0.04	2.847	0.091	−0.06	0.05	1.622	0.202	0.03	0.04	1.07	0.300	**0.50**	**0.04**	**41.876**	**0.000**
TP	**0.55**	**0.07**	**41.513**	**0.000**	**0.20**	**0.06**	**11.573**	**0.000**	**0.22**	**0.06**	**13.456**	**0.000**	**0.14**	**0.06**	**5.664**	**0.017**	**0.20**	**0.05**	**20.042**	**0.000**
Model 3	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*	*β*	*S.E.*	*Δχ2*	*p*
TP	**0.52**	**0.07**	**45.717**	**0.000**	**0.23**	**0.06**	**15.533**	**0.000**	**0.22**	**0.06**	**12.870**	**0.000**	**0.17**	**0.06**	**8.194**	**0.004**	**0.22**	**0.05**	**26.072**	**0.000**
AA	0.08	0.11	0.812	0.367	**0.34**	**0.15**	**5.703**	**0.016**	0.16	0.12	1.667	0.196	0.17	0.14	1.511	0.218	0.09	0.11	0.499	0.479
AE	0.02	0.12	0.350	0.553	−0.17	0.16	1.210	0.271	0.08	0.14	0.662	0.415	−0.09	0.13	0.690	0.405	−0.13	0.11	1.108	0.298
CE	−0.04	0.05	0.431	0.511	**0.12**	**0.06**	**5.023**	**0.025**	0.01	0.05	0.385	0.534	**0.23**	**0.06**	**16.280**	**0.000**	0.05	0.05	1.387	0.238
ACH	0.08	0.05	3.701	0.054	0.04	0.05	1.116	0.290	−0.07	0.05	2.488	0.114	0.01	0.04	0.756	0.384	**0.52**	**0.07**	**39.724**	**0.000**
Gender	**0.29**	**0.08**	**15.773**	**0.000**	−0.17	0.10	3.603	0.057	0.06	0.10	0.451	0.501	−0.10	0.10	1.101	0.294	0.03	0.06	0.646	0.421
Mot. ed.	**0.07**	**0.04**	**3.914**	**0.048**	0.06	0.04	2.207	0.137	0.04	0.04	1.469	0.225	0.05	0.04	1.60	0.184	**0.19**	**0.03**	**35.025**	**0.000**
Classtype	**0.28**	**0.14**	**11.369**	**0.000**	**0.22**	**0.10**	**6.478**	**0.010**	0.15	0.12	1.995	0.157	0.13	0.09	2.335	0.126	**0.19**	**0.09**	**7.271**	**0.007**

Note. The effects of covariates on grade four measures have been included in text within results section. P-values presented in the Table are based on likelihood-ratio tests of model comparison with one degree of freedom. All p-values except for one were consistent with type complex analyses and the results derived from likelihood-ratio tests of model comparisons. Namely, the effect of mother's education on teacher's perceptions at 6th grade was marginally significant in chi-square model comparisons ($\Delta\chi 2 = 3.914$, $p = .048$) whereas in the type complex model it was not ($\beta = 0.07$, $SD = 0.04$, $p = .076$). Therefore, this specific result should be approached with caution and considered as indicative rather than strongly conclusive.

more positive development of agency beliefs of abilities and achievement (see Table 2, Model 3). Specifically, children in classes with a special emphasis showed greater improvements in their ability appraisals and academic performance over the examined period compared to those in general classes.

5. Discussion

This study explored the dynamic interplay between teachers' perceptions, tracking, and the development of children's motivational beliefs and achievements. The significance of teachers' perceptions in shaping children's motivation and achievement has been well-documented (Pesu et al., 2016; Spinath & Spinath, 2005; Upaduay & Eccles, 2014). Similarly, tracking students based on abilities, has been a focal point of educational research due to its potential to either support or hinder the development of students' achievement or motivation, depending on the context and implementation (Chmielewski et al., 2013; Gamoran, 2010). Moreover, the interactions between teachers' perceptions and tracking practices are critical yet underexplored areas. Previous studies have suggested that student grouping can influence teachers' expectations and perceptions of children (e.g., Andersen, 2018; Bohlmann & Weinstein, 2013). Consequently, it has been proposed that this may be one of the ways through which tracking affects students' learning (Pallas et al., 1994). This study sought to provide new insights into these complex interactions, particularly within the unique context of Finnish education, where the curricula across tracks are likely to be more similar than in other tracked contexts, thereby allowing a clearer examination of tracking's effects, less influenced by curricular differences. Our study was organized around three core objectives. First, we sought to investigate how teachers' perceptions influence the development of children's motivational beliefs and achievements. Secondly, we explored how class type, gender, and mother's education level relate to teachers' perceptions of children. Finally, we investigated the role of tracking in influencing changes in children's motivational beliefs and achievements.

Our results confirmed our first hypothesis (H1) and showed that, as we expected based on prior research (e.g., Upaduay & Eccles, 2014), teachers' perceptions of children's academic schoolwork skills predicted changes in children's self-appraisals. Regarding achievement, the relationship between teachers' perceptions and children's achievement seemed to be reciprocal in nature but adding covariates to the model showed that actually it wasn't. Specifically, the level of mother's education significantly influenced the relationship between student achievement and teachers' perceptions. Given that mother's education level was strongly associated with children's test scores at both time points, adding it to the model diminished the previously observed effect of achievement on teachers' perceptions.

Also, our second hypothesis (H2) was supported by our results, indicating that teachers' perceptions differed according to all the examined factors, namely by students' gender, their mothers' education levels, and their class types. In line with previous studies (e.g., Eccles, 2011), our analysis suggests that teachers' perceptions may have a role in shaping the gender differences observed in children's self-appraisals. Teachers perceived girls' academic schoolwork skills more positively than boys' at both time points. Although gender was not directly associated with greater development of motivational beliefs or achievements in our data, prior research suggests that systematic gender differences often exist in school settings (e.g., Early et al., 2020; Voyer & Voyer, 2014). Our findings contribute to this body of literature by confirming the suggestion of previous studies (e.g., Eccles, 2011) that teachers may play a role in reinforcing these differences. Moreover, teachers viewed children with highly educated mothers more favorably, even when controlling for achievement levels. This aligns with research on the role of **socioeconomic status (SES)** and cultural capital in educational attainment (e.g., Early et al., 2020; Jalovaara & Andersson, 2018): children from more educated families often possess higher levels of cultural capital and family resources, which teachers may interpret as a sign of academic ability, consequently rewarding these children with higher grades (Bourdieu & Passeron, 1990/1990; Leopold & Shavit, 2011). Regarding class type, track placement seemed to affect the way children were seen by teachers, as has been suggested before (e.g., Andersen, 2018). In the first years of tracking, children studying in general classes without an emphasis were seen more positively by teachers, but this changed during the following school years, and at sixth grade, children studying in classes with a special emphasis were perceived more positively by teachers at least in terms of their schoolwork skills.

Interestingly, our findings showed positive effects of participation in special emphasis classes in the Finnish comprehensive school system, to our knowledge for the first time. Children in these classes demonstrated greater improvements in achievement and ability beliefs compared to those in classes without an emphasis. As we see effects on the later (sixth-grade) measures of these constructs while controlling for previous (fourth-grade) measures, we assume we are not observing selection effects, but rather effects caused by the tracking itself. This finding contradicted our third hypothesis (H3), in that it is not only differences in curricula between tracks (which are minor in Finland) that seem to cause between-tracks differences in students' development.

Our sample allows us to reduce potential explanations for the more positive developments of students in the special-emphasis track. Not only are there no differences in the core curricula between tracks, but also other instructional variation between tracks should be relatively small in Finland. Prior research has suggested that achievement differences between tracks could also be explained by differences in teachers' experience or qualifications between tracks (Gamoran, 2010; Oakes et al., 1990), but in Finland, however, this should not be the case. Since Finland's education reform in the 1970′s, there has been strong aspiration and national consensus for maintaining the idea that all schools are equal in terms of quality of education (Välijärvi, 2004). Teachers in general all have Master's level of education, and there is no evidence showing any systematic differentiation between schools regarding teachers' experience (Kantasalmi & Kupiainen, 2021). Also, school funding is not dependent on the type of tracks offered, practically all schools are publicly funded, and special tracks are not a criterion in the allocation of funding (Välijärvi, 2004). Basic education is free for all children, regardless of the school or type of the class. Therefore, at the instructional or structural level, the variation between Finnish schools or classes should be relatively small. Thus, we interpret our results as stemming from social-level influences, including the observed effects of teachers, as well as potential influences of class composition and peer interactions (see Pallas et al., 1994).

Our results could be interpreted as stemming from peer influences, which have been extensively studied in educational sciences. More specifically, our results could be interpreted as providing evidence in support of the somewhat contested issue of positive peer influence, specifically the Peer Spillover Effect (Stäbler et al., 2017; Televantou et al., 2021), which suggests that peers' achievements in class may 'spill over,' benefiting individual students.[8] Additionally, the Reflected Glory-Effect (Marsh et al., 2000) could explain our findings of more positive development in children's ability appraisals in classes with a special emphasis. This effect involves comparison processes between 'our class' and other classes within the same school. Although prior

[8] It is important to note that previous research, such as Dicke et al. (2018), has described this as a "phantom effect" that may be falsely detected if appropriate statistical controls and advanced analyses are not conducted. Therefore, our findings contribute to this ongoing debate by suggesting the presence of positive peer influence, while acknowledging the need for careful consideration of potential confounding factors. To elaborate more on the role of peer influence, future studies should explicitly examine peer effects using multilevel models that focus on contextual effects.

research has typically found that the most influential comparisons occur at the most proximal level (i.e., within the class), leading to a more prominent negative effect of the Big-Fish-Little-Pond (BFLP) phenomenon, it has also been shown that self-beliefs are molded through simultaneous comparisons across multiple reference frames (Skaalvik & Skaalvik, 2002; Zell & Alicke, 2010). Additionally, the type of tracking may influence these frames and, consequently, the ways individuals form their self-conceptions (Chmielewski et al., 2013). In Finland, the process of sorting students into classes with and without a special emphasis is not based on ability, and the criteria for selection vary between the emphasized subjects and municipalities (Seppänen, 2003). Nonetheless, based on findings from earlier research (e.g., Berisha & Seppänen, 2017; Kosunen & Seppänen, 2015) and our own data (evidenced by higher maternal education levels and test achievement in our data, see Appendix A), we know that children in classes with a special emphasis typically come from more privileged homes and have higher achievement test scores. While tracking in Finland is not as overt as in more stratified systems, recent Finnish research (Kosunen et al., 2024; Peltola, 2021) has shown that students are acutely aware of these distinctions between classes with and without special emphases, suggesting that they play a role in implicit social comparison processes. Thus, it is reasonable to propose that these distinctions could partially explain the differences we have observed in the development of children's motivation and achievement across different tracks.[9] However, it is important to note that peer influence could only be one explanation for our observed effects, and we did not directly test it in the current research. Building from this understanding, future studies could benefit from employing multilevel SEM to include compositional investigations in their analyses to further explore the effects of tracking. This approach would delve deeper into how the makeup of a class, such as the distribution of students' socio-economic backgrounds and initial academic abilities, influences individual outcomes.

It is worth noting that our modeling approach controlled for most potential confounders. Murayama & Gfrörer (2023) show theoretically that effects of CLPMs can account for many unobserved confounders, with only trait-like confounders affecting both teachers and students challenging the causal interpretability of the cross-lagged effects (e.g., our models cannot account for potential unobserved student-characteristics that influence students' achievement at grade 4 and 6 as well as teachers' assessments of students). Since we controlled for the most reasonable time-stable covariates, (school track, gender, and maternal level of education, seen as a proxy for socio-economic status), our results allow researchers to think clearly about the question whether the observed cross-lagged relationships may be causal in nature, or whether they are due to other time-stable variables not included in our model. Future studies should aim to explore and possibly identify additional trait-like confounders that affect both teachers and students, which may still be influencing the causal interpretations of our findings. Integration of advanced statistical techniques such as propensity score matching could be the next step. This would allow for better control of selection bias, thereby enhancing the robustness of causal claims. Furthermore, conducting detailed moderator analyses for the variables we used as covariates (i.e., gender, mother's education, and class type) could provide deeper insights into whether the impact of teachers' evaluations on children's self-beliefs is similar across various groups and class contexts.

The primary limitation of our study was the limited number of classes, especially those with special emphases, which prevented us from conducting subject-specific comparisons between classes. Addressing this limitation would enable future studies to conduct domain-specific analyses, providing insightful details into the effects of tracking in specific academic areas. For instance, such analyses could examine the development of mathematical thinking skills and math-focused motivational beliefs in classes that emphasize mathematics. We can hypothesize that such a detailed examination would reveal more pronounced effects of tracking than what we observed in our current study, where classes with special emphases were grouped together and measures of competence and motivation were applied more generally. However, in the present study, our data did not permit these types of comparisons between different subjects due to the limited number of classes emphasizing different subjects. Additionally, our study's generalizability across all Finnish municipalities is limited. Practices of establishing classes with special emphases vary significantly between municipalities, with different areas of emphasis and potentially varied consequences for students studying in these environments. Nevertheless, our results can be seen as providing some indication of the effects of studying in classes with special emphases. However, these findings should be corroborated with a larger dataset that would encompass a greater number of classes emphasizing different subjects, thereby enabling subject-specific comparisons.

In our study, the same teachers taught children throughout the study, simplifying the interpretation of results compared to environments where teachers change every year. However, this also limits the generalizability of our findings to contexts where teachers change regularly. Therefore, future research should replicate our study in contexts where teachers change frequently to examine the impact of multiple and different teachers on the development of children's self-beliefs. Additionally, in our setting, children had already studied in classes with a special emphasis for one year before the first assessments and teachers had already been teaching the students for at least one year before the first assessments at fourth grade. This can be viewed either as a strength or as a limitation. In the case of teacher ratings, this can be seen as a strength since it increases the validity of the ratings since teachers already know their students to some extent. However, it may be seen as a limitation for analyses of tracking effects on children's achievement and motivation since children have already studied in different tracks for one year before the first measurement. It is possible that some initial differences in motivation and achievement may have already emerged as a result of this prior involvement in different tracks, potentially confounding our results.

Our study has several practical implications. First, our findings underscore the critical role of teachers' subjective assessments in shaping children's self-perceptions as learners. Given the profound impact that positive motivation can have on individuals' educational attainments and overall well-being (Elliot et al., 2017), it is vital to acknowledge the significant role teachers play in this process. Recognizing this influence highlights the importance of equipping teachers with the necessary tools and knowledge to effectively foster these qualities through their teaching and interactions with students.

Understanding the impact that teachers' have on students' developing self-perceptions is important also because our findings revealed that teachers' perceptions of students' academic schoolwork skills differed based on students' gender, their mothers' education, and class types. For the Finnish educational political discussion, these findings raise some important questions. Finnish basic education system has become internationally acknowledged for its ability to produce top learning results in a relatively equitable manner, with minimal differences between schools and a small achievement gaps between students from different socio-economic backgrounds (e.g, Ahonen, 2021; OECD,

[9] It is important to note that the subject of emphasis may affect these processes. For example, the findings of Peltola (2021), which indicate that students perceive those in special emphasis classes as being different compared to students in general classes and that these differences are often attributed to socioeconomic or racial differences, did not directly address the subject emphasized in the special classes. Nevertheless, the subject of emphasis likely influences how these differentiations affect social comparison processes and, consequently, how students perceive themselves in particular domains. For instance, being in a sports-focused class may have different implications for students' self-beliefs in academic subjects like mathematics than being in a science-focused class.

2023a). Yet, at the same time, gender differences in achievement have been among the largest in the OECD (Leino et al., 2019) as well as the differences between classes within schools (Yang Hansen et al., 2014). Therefore, our findings, which revealed disparities in teachers' perceptions of students' schoolwork skills based on gender, socioeconomic background, and class type, further challenge the notion of an entirely equitable education system. Our results underscore the need for a critical reflection within teacher education programs to better understand and address the factors contributing to these differences in perception. In addition, it highlights the need for further research. It would be important to investigate which factors contribute to and produce these varying perceptions that teachers have about students. One key area for future examination could be the impact of teachers' own backgrounds (e.g., socio-economic background, gender, teaching experience, etc.) on the way they perceive and evaluate students. For example, Lavy (2008) examined gender bias in teachers' grading and found that the bias was explained by teachers' characteristics rather than students' characteristics. In our study, the lack of data on teachers' backgrounds can be seen as a limitation, which we hope future research will address and overcome.

Third, our findings regarding the positive effects of tracking within the Finnish comprehensive school system may have important considerations for educational policy. Tracking students based on their abilities or other aptitudes, either within or between schools, is a common but controversial practice globally. Advocates of tracking see it as a means to ensure that each child receives individualized teaching tailored to their specific needs and abilities (e.g., Tieso, 2010). Opponents, however, highlight the negative impacts on equality and argue that the benefits of tracking are not evenly distributed among students (see Wößmann, 2009). In Finland, the debate for and against tracking has followed these same rationales (Kantasalmi & Kupiainen, 2021; Silvennoinen et al., 2015). Our study does not resolve the ethical dilemma of whether to track or not but enhances our understanding of tracking's effects, grounding the discussion in robust scientific findings. Our findings suggest that tracking may indeed have benefits for students who end up in classes with a special emphasis. However, this does not necessarily mean that our findings endorse tracking. Considering our results, we must assess whether such practices align with the spirit and values of the Finnish comprehensive school system, which was built upon promoting educational equality and equal opportunities for all students, regardless of their backgrounds (Ahonen, 2001). Specifically, it raises an important ethical question: is it fair for those who already possess considerable advantages to receive even more? Ultimately, this decision rests with education policymakers, and therefore, we leave it in their hands.

6. Conclusions

This study examined the impact of teachers' perceptions of students' academic schoolwork skills and tracking practices on students' motivational self-beliefs and achievement. Our findings showed that teachers' perceptions influenced children's motivation and achievement. Notably, these perceptions varied by gender, mother's education level, and class type, with teachers generally viewing girls, children of more educated mothers, and those in classes with a special emphasis more positively than their peers. Furthermore, our analysis demonstrated that being in a class with a special emphasis was associated with more positive developments in both agency beliefs about abilities and actual academic achievement.

Overall, our findings highlight the critical role that teachers play in shaping children's self-perceptions and self-image as learners. Given the observed biases in teachers' perceptions, our results emphasize the need for a critical examination and dismantling of preconceptions within teacher education programs. Additionally, the study underscores the importance of understanding how tracking practices can influence educational outcomes, suggesting a need for policymakers to carefully consider the implementation and potential impacts of such practices.

CRediT authorship contribution statement

Satu Koivuhovi: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Alexander Jung:** Writing – review & editing, Writing – original draft. **Elina Kilpi-Jakonen:** Writing – review & editing. **Todd D. Little:** Methodology. **Mari-Pauliina Vainikainen:** Data curation.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Chat GPT and Grammarly for assistant with fluent academic writing (i.e. grammar and style check). After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This research was supported by the INVEST Research Flagship Centre, University of Turku (funded by Research Council of Finland, decision number: 345546.

Declaration of competing interest

Neither the co-authors nor I have any personal or institutional conflicts of interest that could be construed as influencing the research presented in our manuscript.

Appendix A. Correlations among variables

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. TP grade 6	1												
2. AA grade 6	**0.38**	1											
3. AE grade 6	**0.44**	**0.69**	1										
4. CE grade 6	**0.34**	**0.79**	**0.57**	1									
5. ACH grade 6	**0.43**	**0.31**	**0.15**	**0.25**	1								
6. TP grade 4	**0.64**	**0.27**	**0.27**	**0.22**	**0.35**	1							
7. AA grade 4	**0.22**	**0.34**	**0.30**	**0.27**	**0.15**	**0.29**	1						
8. AE grade 4	**0.28**	**0.25**	**0.33**	**0.22**	**0.08**	**0.41**	**0.85**	1					
9. CE grade 4	**0.18**	**0.30**	**0.19**	**0.35**	**0.21**	**0.25**	**0.54**	**0.49**	1				
10. ACH grade 4	**0.30**	**0.23**	0.04	**0.16**	**0.58**	**0.33**	**0.17**	0.04	**0.21**	1			
11. Class type	**0.13**	**0.14**	0.04	**0.11**	**0.21**	**−0.14**	0.01	−0.02	0.07	**0.09**	1		
12. Gender	**0.38**	**−0.10**	**0.12**	−0.05	0.06	**0.34**	−0.00	**0.21**	−0.02	−0.04	**0.10**	1	

(*continued on next page*)

(*continued*)

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
13. ME	**0.21**	**0.17**	**0.09**	**0.15**	**0.39**	**0.10**	**0.11**	0.07	**0.18**	**0.31**	**0.36**	−0.02	1
M	5.39	5.60	5.30	5.31	21.21	5.44	5.80	5.78	5.07	17.79	–	–	–
SD	1.41	1.11	1.11	1.21	6.93	1.40	1.17	1.10	1.43	6.33	–	–	–
Range	1–7	1–7	1–7	1–7	1–38	1–7	1–7	1–7	1–7	1–34	–	–	–

Note. Abbreviations that are used in the Table are: TP = Teacher's perceptions of children's academic schoolwork skills; AA = Agency beliefs of ability; AY = Agency beliefs of effort; CE = Control Expectancy beliefs; ACH = achievement; ME = maternal education. Statistically significant correlations ($p < .05$) are presented in bold.

Data availability

The authors do not have permission to share data.

References

Ahonen, S. (2001). Kuka tarvitsee yhtenäistä koulua? [Who needs the comprehensive school?]. In A. Jauhiainen, R. Rinne, & J. Tähtinen (Eds.), *Koulutuspolitiikka Suomessa ja ylikansalliset mallit* (pp. 155–184).

Ahonen, A. K. (2021). Finland: Success through equity—the trajectories in PISA performance. In N. Crato (Ed.), *Improving a country's education: PISA 2018 results in 10 countries* (pp. 121–136). Switzerland: Springer.

Andersen, I. G. (2018). Pygmalion in instruction? Tracking, teacher reward structures, and educational inequality. *Social Psychology of Education, 21*(5), 1021–1044.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*, 682–699.

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review, 7*(1), 62–74.

Berisha, A.-K., & Seppänen, P. (2017). Pupil selection segments urban comprehensive schooling in Finland: Composition of school classes in pupils' school performance, gender, and ethnicity. *Scandinavian Journal of Educational Research, 61*(2), 240–254.

Bohlmann, N. L., & Weinstein, R. S. (2013). Classroom context, teacher expectations, and cognitive level: Predicting children's math ability judgments. *Journal of Applied Developmental Psychology, 34*(6), 288–298.

Bond, T. G. (1995). *Blot: Bond's logical Operations test. Townsville, Australia*.

Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education, society and culture* (Vol. 4). Sage.

Butler, R. (2005). Competence assessment, competence, and motivation between early and middle childhood. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 202–221). New York: Guilford Press.

Butler, R. (2008). Ego-involving and frame of reference effects of tracking on elementary school students' motivational orientations and help seeking in math class. *Social Psychology of Education, 11*(1), 5–23.

Butler, R., & Hasenfratz, L. (2017). Gender and competence motivation. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 489–511). New York: Guilford Press.

Carbonaro, W. (2005). Tracking, students' effort, and academic achievement. *Sociology of Education, 78*(1), 27–49.

Chapman, M., & Skinner, E. A. (1989). Children's agency beliefs, cognitive performance, and conceptions of effort and ability: Individual and developmental differences. *Child Development*, 1229–1238.

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.

Chiu, M. M., Chow, B. W.-Y., & Joh, S. W. (2017). Streaming, tracking and reading achievement: A multilevel analysis of students in 40 countries. *Journal of Educational Psychology, 109*(7), 915–934.

Chmielewski, A. K. (2014). An international comparison of achievement inequality in within-and between-school tracking systems. *American Journal of Education, 120*(3), 293–324.

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American Educational Research Journal, 50*(5), 925–957.

Cimpian, A. (2017). Early reasoning about competence is not irrationally optimistic, nor does it stem from inadequate cognitive representations. In A. Elliot, C. Dweck, & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 387–407). New York: The Guilford Press.

Collins, C., & Gan, L. (2013). *Does sorting students improve scores? An analysis of class composition. (No. W18848)*. National Bureau of Economic Research. http://www.nber.org/papers/w18848.

Cooc, N., & Kim, J. S. (2017). Peer influence on children's reading skills: A social network analysis of elementary school classrooms. *Journal of Educational Psychology, 109*(5), 727.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology, 83*(1), 17–27.

Demetriou, A., Pachaury, A., Metallidou, Y., & Kazi, S. (1996). Universals and specificities in the structure and of quantitative-relational thought: A cross-cultural study in Greece and India. *International Journal of Behavioral Development, 19*(2), 255–290.

Demetriou, A., Platsidou, M., Efklides, A., Metallidou, Y., & Shayer, M. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction, 1*(1), 19–43.

Demetriou, A., Spanoudis, G., & Mouyi, A. (2011). Educating the developing mind: Towards an overarching paradigm. *Educational Psychology Review, 23*(4), 601–663.

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *The American Economic Review, 101*(5), 1739–1774.

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review, 52*(2), 245–273.

Early, E., Miller, S., Dunne, L., Thurston, A., & Filiz, M. (2020). The influence of socio-economic background and gender on school attainment in the United Kingdom: A systematic review. *The Review of Education, 8*(1), 120–152.

Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development, 35*(3), 195–201.

Eccles, Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*(3), 830.

Elliot, A., & Hulleman, C. (2017). Achievement goals. In A. Elliot, C. Dweck, & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 43–60). New York: The Guilford Press.

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology, 9*, 1569.

Francis, B., Connolly, P., Archer, L., Hodgen, J., Mazenod, A., Pepper, D., Sloan, S., Taylor, B., Tereshchenko, A., & Travers, M. C. (2017). Attainment grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among year 7 students. *International Journal of Educational Research, 86*, 96–108.

Gamoran, A. (2010). Tracking and inequality. New directions for research and practice. In M. W. Apple, S. J. Ball, & L. A. Gandin (Eds.), *The routledge international handbook of the sociology of education* (pp. 213–228). London: Routledge.

Gentrup, S., & Rjosk, C. (2018). Pygmalion and the gender gap: Do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Educational Research and Evaluation, 24*(3–5), 295–323.

Georgiou, S. N., Christou, C., Stavrinides, P., & Panaoura, G. (2002). Teacher attributions of student failure and teacher behavior toward the failing student. *Psychology in the Schools, 39*(5), 583–595.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every" one" matters. *Psychological Methods, 6*(3), 258.

Gottfried, M. A. (2014). Peer effects in urban schools: Assessing the impact of classroom composition on student achievement. *Educational Policy, 28*(5), 607–647.

Graham, S. (1984). Teacher feelings and student thoughts: An attributional approach to affect in the classroom. *The Elementary School Journal, 85*(1), 91–104.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*(1), 124–136.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*(3–4), 153–166.

Hanushek, E. A., & W ößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*(510), C63–C76.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement (reprinted)*. Routledge.

Hautamäki, J. (1985). Peruskoululaisten loogisen ajattelun mittaamisesta ja esiintymisestä. [The measurement and distribution of Piagetian stages of thinking in Finnish comprehensive school]. *Psykologia, 20*(2), 104–110.

Hautamäki, Arinen, P., Hautamäki, A., Kupiainen, S., Lindblom, B., Niemivirta, M., Pakaslahti, L., Rantanen, P., & Scheinin, P. (2002). *Assessing learningto- learn. A framework*. Helsinki: National Board of Education.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology, 26*(3), 276–295.

Hienonen, N. (2020). Does class placement matter? Students with special educational needs in regular and special classes. *Helsinki studies in education*. University of Helsinki. number 87.

Hoxby, C. (2000). Peer effects in the classroom. Learning from gender and race variation. (No. W7867). In *National bureau of economic research. Cambridge. James cook university. (Original work published 1976)*.

Jalovaara, M., & Andersson, G. (2018). Disparities in children's family experiences by mother's socioeconomic status: The case of Finland. *Population Research and Policy Review, 37*, 751–768.

Jussim, L., Robustelli, S. L., & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 349–380). New York & London: Routledge.

Kantasalmi, K., & Kupiainen, S. (2021). Classes with selective intake in Finnish comprehensive school: A problem of societally equal opportunity for schooling or a boost for learning and equity in pedagogical arrangements? *International Journal of Educational Research, 109*, Article 101857.

Kelly, S., & Carbonaro, W. (2012). Curriculum tracking and teacher expectations: Evidence from discrepant course taking models. *Social Psychology of Education, 15*(3), 271–294.

Kindermann, T. A. (2016). Peer group inlfluences on students' academic motivation. In K. Wentzel, & G. Ramani (Eds.), *Handbook of social influences in school contexts social-emotional, motivation and cognitive outcomes* (pp. 31–46). Routledge.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 8*(5), 363–394.

Koivuhovi, S. (2021). Studying in a class with a special emphasis. In *Changes in children's competence beliefs and mathematical thinking skills and the role of social comparisons*. Helsinki: Unigrafia.

Koivuhovi, S., Vainikainen, M. P., & Kalalahti, M. (2022). The effect of studying in selective classes on the change in pupils' action-control beliefs during lower secondary school in Finland. *Scandinavian Journal of Educational Research, 66*(1), 105–118.

Kosunen, S., Bernelius, V., Seppänen, P., & Porkka, M. (2020). School choice to lower secondary schools and mechanisms of segregation in urban Finland. *Urban Education, 55*(10), 1461–1488.

Kosunen, S., Juvonen, S., & Peltola, M. (2024). *Koulu ja eriarvoisuus*. Helsinki: Gaudeamus [School and inequality].

Kosunen, S., & Seppänen, P. (2015). The transmission of capital and a feel for the game: Upper-class school choice in Finland. *Acta Sociologica, 58*(4), 329–342.

Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72*(5), 1554–1578.

Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., Puhakka, E., Pulkkinen, J., Rautopuro, J., Siren, M., Vainikainen, M.-P., & Vettenranta, J. (2019). PISA 18 ensituloksia. Suomi parhaiden joukossa. Suomalaisnuorten lukutaidon ja lukuharrastuksen muuttuminen. [Initial results of PISA 2018. Finland among the best. Changes in the literacy and reading habits of Finnish youth]. In *Opetus- ja kulttuuriministeriön julkaisuja 2019* (p. 40). Helsinki: Opetus- ja kulttuuriministeriö.

Leopold, L., & Shavit, Y. (2013). Cultural capital does not travel well: Immigrants, natives and achievement in Israeli schools. *European Sociological Review, 29*(3), 450–463.

Little, T. D., Hawley, P., Henrich, C., & Marsland, K. (2002). Three views of the agentic self: A developmental synthesis. In E. L. Deci, & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 177–195). University of Rochester Press.

Little, T. D., Lopez, D. F., Oettingen, G., & Baltes, P. B. (2001). A comparative-longitudinal study of action-control beliefs and school performance: On the role of context. *International Journal of Behavioral Development, 25*(3), 237–245.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81*(3), 417.

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78*(2), 337–349.

Marsh, H. W., & Seaton, M. (2015). The big-fish–little-pond effect, competence self-perceptions, and relativity. In , *Vol. 2. Advances in motivation science* (pp. 127–184). Elsevier.

Marshall, H. H., & Weinstein, R. S. (1984). Classroom factors affecting students' self-evaluations: An interactional model. *Review of Educational Research, 54*(3), 301–325.

Mazenod, A., Francis, B., Archer, L., Hodgen, J., Taylor, B., Tereshchenko, A., & Pepper, D. (2019). Nurturing learning or encouraging dependency? Teacher constructions of students in lower attainment groups in English secondary schools. *Cambridge Journal of Education, 49*(1), 53–68.

Millsapp, R., & Cham, H. (2013). Investigating factorial invariance in longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 109–128). New York: The Guilford Press.

Muenks, K., Wigfield, A., & Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Developmental Review, 48*, 24–39.

Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Järvinen, L. (2012). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *British Journal of Educational Psychology, 82*(2), 185–206.

Murayama, K., & Gfrörer, T. (2022). *Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for model choice from causal inference perspective*.

Muthén, L., & Muthén, B. (2018). *Mplus: Statistical analysis with latent variables: User's guide [version 8]*. Muthén & Muthén. Retrieved [26.5.2021] from: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*(3), 800.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*.

OECD. (2016). In *Equations and inequalities: Making mathematics accessible to all*. Paris: OECD Publishing.

OECD. (2023a). *PISA 2022 results (Volume I and II): Country notes*. Finland. Retrieved [25.09.2024] from https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/finland_6991e849-en.html.

OECD. (2023b). *PISA results 2022 (volume III): Factsheets*. Finland. Retrieved [25.09.2024] from https://www.oecd.org/en/publications/pisa-results-2022-volume-iii-factsheets_041a90f1-en/finland_7c8d87ea-en.html.

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and and school achievement. In R. J. Riding, & S. G. Rayner (Eds.), *International perspectives on individual differences: Vol. 2. Self perception* (pp. 239–265). Ablex Publishing. Retrieved [26.5.2021] from: https://www.uky.edu/~eushe2/Pajares/PajaresSchunk2001.html.

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education, 67*(1), 27.

Parker, Van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction, 54*, 125–137.

Peltola, M. (2021). Everyday consequences of selectiveness. Borderwork in the informal sphere of a lower secondary school in the metropolitan area of Helsinki, Finland. *British Journal of Sociology of Education, 42*(1), 97–112.

Pesu, L., Viljaranta, J., & Aunola, K. (2016). The role of parents' and teachers' beliefs in children's self-concept development. *Journal of Applied Developmental Psychology, 44*, 63–71.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.

Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology, 82*(4), 647–671.

Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society, 26*(2), 290–322.

Riley, T., & Ungerleider, C. (2012). Self-fulfilling prophecy: How teachers' attributions, expectations, and stereotypes influence the learning opportunities afforded Aboriginal students. *Canadian Journal of Education/Revue canadienne de l'éducation, 35*(2), 303–333.

Seppänen, P. (2003). Patterns of 'public-school markets' in the Finnish comprehensive school from a comparative perspective. *Journal of Education Policy, 18*(5), 513–531.

Seppänen, P., Kalalahti, M., Rinne, R., & Simola, H. (2015). Lohkoutuva peruskoulu. Perheiden kouluvalinnat, yhteiskuntaluokat ja koulutuspolitiikka. [Segmenting comprehensive school: Parental school choice, social classes and education policies]. *Finnish Educational Research Association*. ISBN: 978-952-5401-70-7.

Seppänen, P., Pasu, T., & Kosunen, S. (2023). Pupil selection and enrolment in comprehensive schools in urban Finland. In M. Thrupp, P. Seppänen, J. Kauko, & S. Kosunen (Eds.), *Finland's famous education system: Unvarnished insights into Finnish schooling* (pp. 71–86). Singapore: Springer Nature Singapore.

Silvennoinen, H., Kalalahti, M., & Varjo, J. (2015). Why fix something that is not broken? The implementation of school choice policy and parental attitudes towards equality and uniformity of comprehensive school system in Finland. *Athens Journal of Education, 2*(1), 37–52.

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of personality and Social Psychology, 58*(2), 292.

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist, 37*(4), 233–244.

Skinner, E. A., Chapman, M., & Baltes, P. B. (1988). Control, means-ends, and agency beliefs: A new conceptualization and its measurement during childhood. *Journal of Personality and Social Psychology, 54*(1), 117–133.

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology, 82*(1), 22–32.

Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development, 63*(2/3).

Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development, 20*(2), 190–204.

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology, 109*(2), 191.

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research, 86*(4), 849–899.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*(2), 213–226.

Televantou, I., Marsh, H. W., Dicke, T., & Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement: Evidence from English early primary schools. *Learning and Instruction, 71*, Article 101399.

TENK. (2012). Responsible conduct of research and procedures for handling allegations of misconduct in Finland. In *Guidelines of the Finnish advisory board on research*

integrity 2012. Finnish Advisory Board on Research Integrity TENK. Retrieved [26.5.2021] from: https://www.tenk.fi/en/tenk-guidelines.
Terrin, É., & Triventi, M. (2023). The effect of school tracking on student achievement and inequality: A meta-analysis. *Review of Educational Research, 93*(2), 236–274.
Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review, 26*(1), 29–36.
Timmermans, A. C., de Boer, H., & van der Werf, M. P. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education, 19*, 217–240.
Trautwein, U., Köller, O., Lüdtke, O., & Baumert, J. (2005). Student tracking and the powerful effects of opt-in courses on self-concept: Reflected-glory effects do exist after all. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *New frontiers for self research* (pp. 307–327). Greenwich, Ct: Information Age Publishing.
Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788.
Upadyaya, K., & Eccles, J. S. (2014). How do teachers' beliefs predict children's interest in math from kindergarten to sixth grade? *Merrill-Palmer Quarterly, 60*(4), 403–430.
Upadyaya, K., & Eccles, J. (2015). Do teachers' perceptions of children's math and reading related ability and effort predict children's self-concept of ability in math and reading? *Educational Psychology, 35*(1), 110–127.
Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology, 44*(5), 331–349.
Välijärvi, J. (2004). The system and how does it work: Some curricular and pedagogical characteristics of the Finnish comprehensive school. *Education Journal, 32*(1), 31–55.
Van Houtte, M., Demanet, J., & Stevens, P. A. (2013). Curriculum tracking and teacher evaluations of individual students: Selection, adjustment or labeling? *Social Psychology of Education, 16*, 329–352.
Varjo, J., & Kalalahti, M. (2019). The art of governing local education markets – municipalities and school choice in Finland. *Education Inquiry, 10*(2), 151–165.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620
Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation, 24*(3–5), 124–179.
Wechsler, D. (1981). *WAIS-R: Manual: Wechsler adult intelligence scale – revised*. Harcourt Brace Jovanovich for Psychological Corp.
Weinstein, R. (2002). *Reaching higher: The power of expectations in schooling*. Cambridge, MA: Harvard University Press.
Wentzel, K. R., & Muenks, K. (2016). Peer influence on students' motivation, academic achievement, and social behavior. In *Handbook of social influences in school contexts* (pp. 13–30). Routledge.
Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x
Wößmann, L. (2009). International evidence on school tracking: A review. https://www.ifo.de/DocDL/dicereport109-rr1.pdf.
Yang Hansen, K., Gustafsson, J., & Rosen, M. (2014). School performance differences and policy variations in Finland, Norway and Sweden. In T. Hallvard (Ed.), *Northern lights on TIMSS and PIRLS 2011: Differences and similarities in the nordic countries* (pp. 25–48). Nordic Council of Ministers. Retrieved [25.9.2024] from https://www.uv.uio.no/ils/forskning/prosjekter/timss/publikasjoner/northernlightstimss-pirls.pdf.
Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review, 14*(4), 368–384.
Zimmer, R. (2003). A new twist in the educational tracking debate. *Economics of Education Review, 22*(3), 307–315.