

A systematic review of the impact of ability grouping on achieving SDG4 in mathematics education

Quality Education
for All

441

Telma Largent

*Department of Communication, Quality Management and Information Systems,
Mid Sweden University, Östersund, Sweden*

Received 27 December 2024
Revised 22 March 2025
Accepted 24 May 2025

Abstract

Purpose – Mathematics education is a gateway to careers in science, technology, engineering and mathematics (STEM), which are essential for sustainable development. The purpose of this study is to analyze existing literature to determine whether ability grouping in mathematics education functions as a driving or restraining force in achieving inclusion, equity and quality education in alignment with Sustainable Development Goal 4 (SDG4).

Design/methodology/approach – A systematic literature review of ability grouping in Grades K–8 (Years 1–9) was conducted. Lewin’s Force Field Analysis was used to systematically evaluate conflicting research findings and determine whether ability grouping advances or hinders SDG4 objectives.

Findings – The findings of this study suggest ability grouping restrains SDG4 progress. While its impact on academic achievement remains debated, concerns persist over placement bias and negative effects on students’ self-concept and growth mindset. To fully align with SDG4’s holistic vision, schools must ensure that student grouping practices positively support each of the SDG4 components: inclusion, equity and quality.

Originality/value – This study offers new insights by applying Force Field Analysis to synthesize conflicting research, providing a structured approach for evaluating the impact of ability grouping on SDG4.

Keywords Ability grouping, Mathematics education, SDG4, Equity, Inclusion, Quality education

Paper type Research paper

Introduction

Education plays a pivotal role in shaping individual opportunities and driving global development. The United Nations Educational, Scientific and Cultural Organization (UNESCO) recognizes education as a critical pillar for sustainable development (Snyder, 2023; UNESCO, 2015). At the 2015 World Education Forum, participants shared a vision of transforming lives through education, emphasizing its essential role in achieving all United Nations Sustainable Development Goals (SDGs) (UNESCO, 2015). Among these, Sustainable Development Goal 4 (SDG4) focuses on inclusive and equitable quality education and lifelong learning opportunities for all (UNESCO, 2015). SDG4 targets, such as ensuring youth literacy, numeracy and equitable, quality primary and secondary education, are crucial to enable upward mobility and reduce workforce inequalities (UNESCO, 2015).



© Telma Largent. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Quality Education for All
Vol. 2 No. 1, 2025
pp. 441-461
Emerald Publishing Limited
2976-9310
DOI 10.1108/QEA-12-2024-0154

Quality education is a goal and the foundation for sustainable development across life domains (Jamali *et al.*, 2022), fostered by the knowledge and skills necessary for future work. Among the skill sets, science, technology, engineering and mathematics (STEM) are identified as essential for achieving SDG targets. STEM disciplines drive innovation across sectors (Jamali *et al.*, 2022; NSF, 2020) through diverse perspectives and talents (Douglas and Attewell, 2017; NSF, 2020). However, many countries face persistent challenges in ensuring equitable access, resulting in a STEM talent crisis (Dahlberg, 2021; National Science Board, 2024). Inequitable practices in mathematics education may contribute to possible STEM workforce shortages. Mathematics education is recognized as a gateway to STEM (Boaler, 2024; Douglas and Attewell, 2017; NCTM, 2018; Redmond-Sanogo *et al.*, 2016). Ability grouping in mathematics may contribute to barriers to STEM access by affecting students' beliefs about their ability to succeed. Boaler (2024) emphasizes the role of key success factors in mathematics, such as a growth mindset. While ability grouping is intended to address students' diverse learning needs (Gupta *et al.*, 2023; Sukhnandan and Lee, 1998), some research suggests that these practices often reinforce inequities and limit opportunities for many students (Boaler, 2015; Archer *et al.*, 2018).

Ability grouping organizes students based on perceived ability or prior performance to create more homogeneous learning environments (Boaler, 2015; Duflo *et al.*, 2011). Terminology varies across educational systems, with terms such as tracking, streaming, setting and sorting often used interchangeably (Boaler, 2015; Steenbergen-Hu *et al.*, 2016). Ability grouping can take different forms. One example is between-class grouping, which Gamoran and Hallinan (1995) describe as "tracking," where students are placed in separate classes for one or more subjects based on ability (Steenbergen-Hu *et al.*, 2016). Another form is within-class grouping, where students are divided into smaller ability-based groups within the same classroom (Sukhnandan and Lee, 1998). A third type is cluster grouping, which places high-achieving students alongside mixed-ability peers to facilitate differentiated instruction (Steenbergen-Hu *et al.*, 2016). Besides the grouping within the schools, in some countries, particularly parts of Europe and Asia, between-school tracking is more pronounced, with students assigned to separate programs – often in different school buildings – preparing them for university or vocational careers (Chmielewski, 2014).

The extent to which ability grouping is used depends on national educational philosophies. Countries like the USA and the UK widely implement ability grouping. In contrast, high-performing systems such as Finland, Japan and Korea prioritize mixed-ability learning environments emphasizing equity and shared learning (Boaler, 2020). For younger students, the within-class grouping may be more prevalent (Boliver and Capsada-Munsech, 2021), but some schools also use between-class grouping at this stage (Boaler, 2020; Slavin, 1987). Von Hippel and Cañedo (2022) point out that regardless of the grouping method, introducing ability grouping early in a child's education – such as kindergarten – often leads to long-term patterns that continue into later grades. Hanushek and Woessmann (2006) highlight that early tracking increases educational inequality. Over time, these groupings can evolve into more rigid tracking systems, shaping students' academic trajectories (Von Hippel and Cañedo, 2022).

Mathematics is often considered well-suited for ability grouping because of its structured, sequential nature, which is perceived as requiring differentiated pacing for students at different skill levels (Sukhnandan and Lee, 1998). However, research on ability grouping yields mixed findings (Legette and Kurtz-Costes, 2020); some studies suggest that grouping can enhance learning by tailoring instruction to student's skill levels (Steenbergen-Hu *et al.*, 2016), benefiting high-ability students by offering appropriately challenging coursework (Boliver and Capsada-Munsech, 2021). Others indicate that it reinforces inequalities and

affects students' long-term academic trajectories (Legette and Kurtz-Costes, 2020). Critics contend that labeling students based on perceived abilities undermines high-quality mathematics education goals and limits opportunities for deep understanding, critical thinking and engagement in mathematics (NCTM, 2023a), thus serving as a gatekeeper to STEM careers (Boaler, 2024; Douglas and Attewell, 2017; Redmond-Sanogo *et al.*, 2016).

Whether ability grouping aligns with SDG4's goal remains an open question. Understanding the structural barriers to achieving SDG4 requires a multi-layered perspective on education systems. Boeren (2019) conceptualizes education through three levels: the micro- (individual learners and their psychological and socioeconomic backgrounds), the meso- (training environments and school policies) and the macro-level (national education policies and systemic structures). While data and reports on SDG4 attainment exist at the micro- (e.g. PISA) and the macro-level (e.g. European Commission and OECD reports), there is a gap in the meso-level data and reporting (Boeren, 2019). The meso level, which translates broad policies into classroom practice (Boeren, 2019), is particularly relevant for evaluating practices that may impact the achievement of SDG4.

In this study, ability grouping is defined as a meso-level teaching practice. Although a substantial body of research exists – some addressing academic achievement, others non-cognitive outcomes and some both – these findings are seldom considered in relation to the achievement of SDG4. Given the widespread use of ability grouping in mathematics education (Boliver and Capsada-Munsech, 2021; Gupta *et al.*, 2023), further research is needed to explicitly examine its impact on inclusion, equity and quality as they relate to achieving SDG4.

This study addresses this gap using a systematic literature review and Force Field Analysis to examine whether ability grouping acts as a driving force promoting SDG4 objectives or as a restraining force obstructing SDG4 attainment. The following research question guides the analysis:

- RQ1. What insights does existing literature provide regarding the role of ability grouping in mathematics education as a driving or restraining force in advancing inclusion, equity and quality education in alignment with SDG4?

Analytical framework

This study applies Kurt Lewin's Change Theory to examine how ability grouping functions within educational systems. Lewin's (1951) three-stage model – unfreezing, moving and refreezing – offers a structured approach to understanding systemic change. "Unfreezing" involves identifying existing practices and recognizing the forces facilitating or resisting change. "Moving" represents transitioning to new educational practices, while "refreezing" ensures that changes become institutionalized through policies and reinforcement. Force Field Analysis, a key component of Lewin's model, assesses the balance between driving forces (which promote change) and restraining forces (which hinder it) (Cameron and Green, 2024). In this study, Force Field Analysis is used to "unfreeze" the current state of mathematics education by identifying whether ability grouping is a practice that advances inclusion, equity and quality.

Sustainable Development Goal 4 (SDG4) is a framework for analyzing the need for inclusive, equitable and quality education for all (UNESCO, 2020). The first key dimension, inclusion, relates to students' non-cognitive outcomes, such as growth mindset, self-concept, enjoyment and motivation. Non-cognitive skills are personality traits, thought patterns, feelings and behavior (Borghans *et al.*, 2008). While none of the reviewed studies explicitly linked non-cognitive outcomes to inclusion within SDG4, these elements reflect students'

emotional and psychological experiences, which are strongly influenced by their sense of belonging (Allen *et al.*, 2024). A strong sense of belonging is critical for students' psychosocial well-being and academic success (Chiu *et al.*, 2016). In contrast, exclusion can lead to lower motivation, reduced enjoyment of learning (Goodenow and Grady, 1993) and diminished self-concept and growth mindset (Dweck *et al.*, 2014). UNESCO (2020) defines inclusion as ensuring that every learner feels valued and respected and experiences a strong sense of belonging within their school environment. This study categorizes non-cognitive outcomes under inclusion, recognizing that fostering a sense of belonging is integral to SDG4's goal of ensuring inclusive and equitable education for all learners.

Another core element of SDG4 is equity, which focuses on ensuring fair access to learning opportunities (UNESCO, 2017). However, as Levinson *et al.* (2022) highlight, educational equity lacks a universally accepted definition and may be interpreted in multiple ways: equal distributions of outcomes across populations; equal outcomes for every child; equal resource allocations across students, schools, districts, states or nations; equal experiences for each child; and equal levels of growth by each child. Although these perspectives vary, these interpretations are often linked to broader goals, such as supporting disadvantaged groups, achieving educational adequacy or balancing short-term benefits with long-term systemic change (Levinson *et al.*, 2022). For this study, research examining placement bias or the disproportionate impact of ability grouping on marginalized student groups is categorized under equity, regardless of the specific definition used in the original studies.

Finally, quality – the third component of SDG4 – is frequently assessed through academic achievement, often measured by standardized assessments. However, education researchers emphasize that quality education extends beyond test scores, including critical thinking, problem-solving and depth of understanding (NCTM, 2023b). This distinction between breadth and depth is vital. Seeley (2004) clarifies that “hard arithmetic is not deep mathematics.” Depth does not mean mastering arithmetic at an earlier age or practicing calculations with more digits. Instead, it involves balancing arithmetic with other essential areas like measurement, geometry and data analysis, fostering problem-solving skills and critical exploration over rote practice (Seeley, 2004).

While standardized assessments remain a key indicator, SDG4 and broader educational policies advocate for multiple measures of learning that offer a more comprehensive view of student development (UNESCO, 2020). For this study, research on academic achievement is categorized under “quality,” recognizing the limitations of standardized assessments in capturing the full scope of educational outcomes.

The three components of SDG4 – inclusion, equity and quality – are shaped by systemic forces that either support or hinder their achievement. Conflicting findings from ability grouping studies (Legette and Kurtz-Costes, 2020) make it difficult for mathematics educators and leaders to make evidence-based decisions. Force Field Analysis helps address this by systematically assessing whether ability grouping supports or obstructs the attainment of SDG4 objectives.

Methods

A systematic literature review was conducted following Creswell and Creswell's (2018) guidelines, using the SCOPUS and ERIC databases and snowballing techniques. The initial search in SCOPUS used the keywords “ability AND grouping AND math,” with filters applied for English-language publications from 2016 to 2024. SCOPUS does not provide specific filtering for peer-reviewed articles. A subsequent search was conducted in ERIC, using the exact keywords and filters but limited results to peer-reviewed, English-language articles. Titles and abstracts were screened for relevance, and full-text reviews were conducted for

selected articles. Duplicates identified between the databases were removed. Additional studies identified through snowballing and expert recommendations were included.

Selection criteria

The review included studies on ability grouping in Grades K–8, equivalent to Years 1–9 (the UK). Studies involving any form of mathematics ability grouping were considered. While grouping methods differ, this study is not focused on the individual impacts of each type. Instead, it investigates the broader effects of labeling students – whether explicitly or implicitly – as low, medium or high achievers when they are grouped homogeneously.

Single studies and meta-analyses were included if they reported cognitive (e.g. academic achievement) or non-cognitive (e.g. growth mindset and self-concept) outcomes in mathematics. For studies covering multiple subjects, only mathematics-related findings were considered. If results were aggregated (e.g. math and English), then the impact was assumed to be the same for both. Studies focusing solely on non-mathematics subjects were excluded.

Data sets

The SCOPUS search yielded 38 articles (Table 1). After screening titles and abstracts, 19 were selected for full-text review, with 11 deemed relevant. The ERIC search produced 20 articles; seven were thoroughly reviewed after screening; and one was considered relevant. An additional seven studies were identified through snowballing. In total, 17 studies (11 from SCOPUS, 1 from ERIC and 5 from snowballing) were included in the force analysis.

Data analysis

Two-step analysis: Thematic grouping and force field analysis

Data analysis occurred in two phases: first, thematic grouping categorized articles by SDG4 components (inclusion, equity and quality); and second, partial force field analysis identified factors driving or restraining ability grouping’s impact on these components in mathematics education.

Thematic grouping

Studies were coded based on their primary focus: inclusion, equity and quality, as previously defined in the analytical framework. Studies addressing academic achievement were coded as quality. Studies examining placement bias or the disproportionate impact of ability grouping on specific student groups were coded as equity. Studies examining non-cognitive outcomes were coded as inclusion.

Table 1. Database search progress

	SCOPUS	Eric
Keywords	Ability AND grouping AND math; English; 2016–2024	Ability AND grouping AND math; English; 2016–2024; peer-reviewed
Identification	<i>n</i> = 38	<i>n</i> = 20
Screening: Read title and abstract	<i>n</i> = 19	<i>n</i> = 7 (after removing seven duplicates from SCOPUS search)
Included: after reading article	<i>n</i> = 11	<i>n</i> = 1
Source(s): Author’s own work		

In addition to grouping studies by focus, each study was coded as positive, negative, neutral or mixed, following these guidelines:

- *Positive*: This study indicated that ability grouping was beneficial for all student groups or at least one group, while it was neutral for others.
- *Neutral*: This study showed no significant positive or negative effects for any group.
- *Negative*: This study indicated negative outcomes for all student groups or detrimental outcomes for at least one group, while it was neutral for others.
- *Mixed*: This study showed positive outcomes for some groups and negative for others.

Force field analysis: Identifying and assessing forces

To examine how ability grouping may advance or hinder progress toward SDG4, a force field analysis technique based on Lewin's Change theory (Snyder and Anderson, 1986) was used to evaluate whether ability grouping serves as a driving or restraining force in achieving SDG4. The first four steps of the force field analysis were followed, identifying and assessing these forces without implementing strategies to amplify or mitigate them:

- (1) identify the problem: identify what role does ability grouping in mathematics education play in helping schools achieve the SDG4 goals;
- (2) identify the various forces at play in the field: examine the literature review studies within the context of SDG4;
- (3) assign a positive or negative label to each force (Table 3); and
- (4) assign a value to each label to indicate its strength (Table 3):
 - rank the forces under the headings "restraining" and "driving" (Tables 6, 7 and 8);
 - identify actions to change field dynamics: not applicable to this study; and
 - develop an action plan for change: not applicable to this study.

Points system for assessing the studies

A points-based system was developed to evaluate the strength and relevance of each study (Table 2). Given the variation in study design, direct comparison can be challenging. Points were assigned across six key categories: data source, type of study, data recency, time horizon, study design and sample size. The points system reflects each category's relative impact of different design and methodological choices. For instance, in the data-source category, a 0.5-point difference between studies using primary and secondary sources versus only primary sources reflects a moderate increase in data comprehensiveness. In the type-of-study category, a 1.5-point difference between studies using quantitative and qualitative methods versus a single method underscores the added depth provided by mixed methods.

Methodology for assigning points

The point system was designed to reflect the robustness and reliability of the study findings. While various research approaches provided valuable insights, greater weight was assigned to specific approaches based on their potential to address complex issues in mathematics education:

- *Data source*: Studies incorporating primary and secondary data received more points than those relying on a single source. Primary data were prioritized for studies with a single data type, as they are typically more directly aligned with the study aims.

Table 2. Points system to evaluate the study design and methodology

Category	Criteria	Points	Quality Education for All
Data source	Primary and secondary	3	447
	Primary only	2.5	
Type of study	Secondary only	2	
	Quantitative and qualitative	3	
Data recency	Quantitative or qualitative	1.5	
	<5 years	3	
	5–10 years	2.5	
Time horizon	>10 years	1	
	Not specified	0	
	Longitudinal	3	
Study design	Cross-sectional	1.5	
	Second-order meta-analysis	5	
	Meta-analysis	4	
	Experimental (RCT)	4	
Sample size	Quasi-experimental	3.5	
	Non-experimental	2	
	Meta-analysis > 10 math studies. Single study quantitative $n > 350$ based on ESSA; qualitative based on Creswell and Creswell's guidelines	3	
	Meta-analysis < 10 math studies. Single study quantitative $n < 350$ based on ESSA; qualitative based on Creswell and Creswell's guidelines	1	
	Not specified	0	

Source(s): Author's own work

- *Type of study:* Studies combining quantitative and qualitative methods received more points for their comprehensive approach. As [Creswell \(2021\)](#) argues, integrating both methods provides a deeper understanding of research problems than using any approach alone.
- *Data recency:* Studies that used more recent data received higher points because they were likely to reflect current educational policies and practices.
- *Time horizon:* Longitudinal studies were awarded more points because of their ability to track changes over time, facilitating the examination of causal relationships in ability grouping. Cross-sectional studies, while helpful, offer only a single snapshot in time.
- *Study design:* This category was prioritized because the study aimed to assess causality between ability grouping practices and students' cognitive and non-cognitive skills. Second-order meta-analyses and meta-analyses were assigned higher starting points (5 and 4, respectively) because of their comprehensive nature. For single studies, points were based on the evidence tiers outlined by the Every Student Succeeds Act (ESSA) [[Institute of Education Sciences \(IES\), 2025](#)]. Randomized controlled trials were valued highly for their ability to establish causality, as random assignment minimizes bias and enhances group comparability. Quasi-experimental studies, while valuable, lack randomization, making their conclusions less definitive. Correlational studies, which provide insights into variable relationships but lack causality, were classified as non-experimental.

- *Sample size:* For quantitative studies, [Creswell and Creswell \(2018\)](#) suggest basing sample size on an analysis plan rather than population fractions or previous studies. Following ESSA guidelines, higher points were assigned to quantitative studies with samples of over 350 students. The appropriate sample size in qualitative research varies by study type: 1–2 participants in narrative studies, 3–10 in phenomenology, 4–5 in case studies and 20–30 in grounded theory ([Creswell and Creswell, 2018](#)).

Total-points calculation

After assigning points to each study, totals were calculated to indicate positive, negative, neutral and mixed impacts. For each category (inclusion, equity and quality), a weighted average was computed by subtracting restraining (negative) forces from driving (positive) ones and dividing them by the total of all forces (positive, negative, neutral and mixed). Finally, the weighted average of all combined forces was calculated.

Results

The systematic literature review findings are organized into two sections. First, the thematic analysis is presented, where articles are coded and grouped by the three main themes aligned with SDG4: quality (measured by academic achievement), equity and inclusion ([Table 3](#)). A narrative summary of each study's findings is then provided and organized by these themes. Section 2 presents partial force field analysis results based on the point system.

Thematic analysis

[Table 3](#) summarizes the studies in the systematic review, detailing publication year, school level, country of study, type of ability grouping and the central SDG4 theme addressed (quality, equity or inclusion). To ensure clarity, ability grouping labels were standardized based on equivalencies across studies, as terminology varied considerably. Depending on the study, grouping types were classified as between-class, within-class or both. One study was categorized as within-school/between-school, as the authors indicated that both types were included in their studies. Where the type of ability grouping was not explicitly stated, the label "ability grouping" was applied. The table also categorizes each study's findings as positive, negative, neutral or mixed in relation to the impact of ability grouping. Studies involving students aged 11–14 years were coded as secondary school, while those involving students under 11 years were coded as primary.

Of the 17 studies, 7 focused on academic achievement and were categorized under quality. Three studies addressed placement bias or equity issues and were classified under equity, while six on non-cognitive skills were classified under inclusion. One study addressed quality and inclusion, resulting in multiple classifications. Of the 17 studies, 5 were at the primary school level, 7 at the secondary level and 5 spanned both primary and secondary levels. Most studies focused on Western countries. Furthermore, the ability grouping type varied, with 13 studies examining a format equivalent to between-class grouping.

Narrative summary of each study's findings

Theme 1: Impact of ability grouping on quality. Under the quality theme, academic achievement, as measured by assessments, served as the primary metric for evaluating quality in mathematics education. Several studies highlighted the potential benefits of ability grouping in improving academic achievement for students across distinct levels.

[Collins and Gan \(2013\)](#) found that reducing the range of ability levels within classrooms through grouping improved student achievement outcomes. Similarly, [Duflo et al. \(2011\)](#)

Table 3. Studies grouped by thematic themes: quality, equity and inclusion

Author	Year	School	Country	Type of ability grouping	Main focus	Impact
Collins and Gan	2013	Primary	The USA: Dallas	Between-Class	Quality	Positive
Dufo <i>et al.</i>	2011	Primary	Kenya	Between-Class	Quality	Positive
Pierce <i>et al.</i>	2011	Primary	The USA	Within-Class: Gifted	Quality	Positive
Sorensen <i>et al.</i>	2017	Primary and secondary	The USA: North Carolina	Between-Class	Quality	Positive
Boaler and foster	2021	Secondary	The USA	Between-Class	Quality	Negative
Burris <i>et al.</i>	2006	Secondary	The USA: Long Island	Between-Class	Quality	Negative
Deunk <i>et al.</i>	2018	Primary	The USA, the UK and Australia	Between, within-Class	Quality and inclusion	Negative
Gupta <i>et al.</i>	2023	Secondary	Rural China	Between-Class	Quality and inclusion	Quality: Negative inclusion: Positive
Archer <i>et al.</i>	2018	Secondary	England	Between-Class	Equity	Negative
Hartas	2018	Secondary	England and Wales	Between-Class	Equity	Negative
Von Hippel and Cañedo	2022	Primary	The USA	Within-Class	Equity	Negative
Bolliver and Capsada-Munsech	2021	Primary and secondary	The UK	Between and within-Class	Inclusion	Negative
Campbell	2021	Primary and secondary	England	Ability grouping	Inclusion	Negative
Francombe and Hewitt	2018	Secondary	England	Between-Class	Inclusion	Negative
McDool	2019	Primary and secondary	The UK	Between-Class	Inclusion	Negative
Legette and Kurtz-Costes	2020	Secondary	The USA	Between-Class	Inclusion	Mixed
Parker <i>et al.</i>	2021	Primary and secondary	28 countries: OECD in 2003	Within-School between-School	Inclusion	Mixed

Note(s): OECD: Organization for Economic Co-operation and Development; the UK: the United Kingdom; the USA: the United States

Source(s): Author's own work

conducted an experimental study that provided evidence of a positive and significant impact on mathematics scores across different student groups, with these benefits persisting over time. [Sorensen et al. \(2017\)](#) found that the effects of peer academic achievement varied by grade level. In fourth grade, a diverse range of academic abilities among peers did not negatively impact students' mathematics performance. However, by seventh grade, increased variance in peer achievement began to have a detrimental effect, suggesting that homogeneous groups are more beneficial at this stage. [Pierce et al. \(2011\)](#) demonstrated that cluster grouping for gifted students led to significant academic gains without negatively affecting non-gifted peers' performance.

In contrast, other studies revealed the adverse effects of ability grouping, particularly for lower-ability students or students from disadvantaged backgrounds. [Boaler and Foster \(2021\)](#) concluded that de-tracking and high-quality professional development can significantly improve student achievement and promote equity. [Burris et al. \(2006\)](#) argued against ability grouping, finding that an accelerated mathematics curriculum in heterogeneously grouped classes led to significant improvements for all student groups, including minority and low socioeconomic status students and those at all achievement levels. Importantly, their study demonstrates that high achievers are not disadvantaged by being in a mixed-ability setting.

Finally, [Deunk et al. \(2018\)](#), in their meta-analysis covering five studies on ability grouping in mathematics, found that the impact of ability grouping on low-ability students was negative in the context of between-class homogeneous grouping. The effect size of -0.300 was considered small but statistically significant, suggesting that while the impact may not be large, it was meaningful. This effect was observed across multiple subjects, including mathematics. The effects were generally neutral for average- and high-ability students ([Deunk et al., 2018](#)). [Gupta et al. \(2023\)](#) found no significant overall impact of ability grouping on academic or non-academic outcomes for high- or low-ability students. However, it negatively impacted low-ability boarding students' academic achievement.

Theme 2: Impact of ability grouping on equity. Studies focusing on placement bias or the impact of ability grouping on marginalized student groups were categorized under equity, reflecting SDG4's emphasis on ensuring equitable access to quality education. For example, [Archer et al. \(2018\)](#) found that privileged students, specifically White, middle-class children, were far more likely to be placed in top-ability groups, while working-class and Black students were often placed in lower-ability groups. Those in the lowest groups, including those receiving free school meals, were particularly critical of the fairness of the grouping system. In contrast, students in the top groups viewed their placement as just. This division in perception highlights how ability grouping reflects and perpetuates social inequalities. Similarly, [Hartas \(2018\)](#) demonstrated that teacher perceptions played a decisive role in grouping decisions. Students who were perceived to exhibit negative attitudes, behaviors or low aspirations were more often placed in middle or lower-ability groups, with this bias disproportionately affecting boys and children from lower-income families. The study also found that the future expectations of 11-year-olds strongly aligned with their assigned positions, leading some to view these placements as fixed.

[Von Hippel and Cañedo \(2022\)](#) found that while test scores were the primary factor in kindergarten ability-group placement, social biases favored girls, high-socioeconomic status children and Asian Americans. By spring, many high-socioeconomic-status children were placed in higher groups than their score gains warranted. Teacher-reported behaviors explained some of the higher placements for girls. Still, they did not account for the elevated placements of high socioeconomic status and Asian American students, highlighting the influence of social bias in perpetuating inequities in ability grouping.

Theme 3: Impact of ability grouping on inclusion. Students' sense of belonging at school is central to their psychosocial well-being and academic success (Chiu *et al.*, 2016). Non-cognitive skills, such as growth mindset, self-concept, motivation and enjoyment, significantly influence students' educational experiences. Students who feel excluded may become less motivated, experience reduced academic enjoyment (Goodnew and Grady, 1993) and develop lower self-concepts and growth mindset (Dweck *et al.*, 2014).

Research highlighted both positive and negative effects of ability grouping on inclusion. Some studies indicated that ability grouping can have positive effects. Gupta *et al.* (2023) found no significant overall impact of ability grouping on students' self-concept for either high- or low-ability students; however, they noted that ability grouping reduced mathematics anxiety among high-ability students.

Conversely, several studies reported adverse outcomes associated with ability grouping. Boliver and Capsada-Munsech (2021) suggested that being placed in a lower-ability group at the age of 7 years reduces the likelihood of a student developing, maintaining or increasing enjoyment in mathematics by age 11 years, even when controlling for mathematics ability, gender and social class at age 7 years. Campbell (2021) found that children in the lowest groups were likelier to develop negative mathematics self-concept later. While boys in the highest-ability groups were unlikely to experience negative self-concept, low-scoring girls in the highest groups were more prone to negative self-concept mainly when influenced by negative teacher judgments.

McDool (2019) observed that ability grouping could negatively affect non-cognitive skills, such as emotional and peer skills, especially for boys, who exhibited more internalizing behaviors. However, being in the lowest group did not significantly harm non-cognitive outcomes for either boys or girls. The study advised caution when transitioning from heterogeneous to ability grouping. Francome and Hewitt (2018) noted that mixed-ability grouping fostered stronger growth mindsets and collaborative learning, whereas homogeneous ability grouping emphasized procedural tasks, limiting opportunities for exploration.

Legette and Kurtz-Costes (2020) reported that students in honors mathematics classes had higher self-concepts than those in regular courses, with classroom appraisals and perceptions of teacher expectations playing a significant role in this disparity. On the other hand, Parker *et al.* (2021) found that ability grouping affects academic self-concept differently across student groups. It may lower self-concept for advantaged students, leading to less ambitious paths, while boosting it for disadvantaged students, though rigid educational tracks still limit their opportunities. The Big-Fish-Little-Pond Effect (BFLPE) enhances math self-concept in highly stratified systems, as students feel more confident among lower-achieving peers. This effect arises because students often evaluate their abilities relative to others in their local context rather than based on their actual abilities. The BFLPE tends to be stronger at class than school levels (Marsh *et al.*, 2014). However, Parker *et al.* also identify a "Perverse Robin Hood Effect," where disadvantaged students gain confidence in lower-performing groups but face systemic barriers that prevent real educational or career advancement. As a result, working-class students develop inflated self-concepts compared to similarly skilled upper-class peers but remain steered toward less ambitious paths because of stratification's signaling effects.

Partial force field analysis: Assessing force results

Force assessment: The force field analysis was derived from the point-allocation system (Table 2) based on data provided in Tables 4 and 5. Table 4 includes information for the studies that addressed quality, and Table 5 provides information for the studies that addressed equity or inclusion. Table 4 indicates that four studies reported a negative impact of ability

Table 4. Study design, methodology and the impact of ability grouping on quality

Author publication year	Data source	Type of data	Data recency	Time horizon	Study design	Sample size	Focus and force
Collins and Gan (2013)	Secondary	Quantitative	2003–2005	Longitudinal	Quasi-experimental	$n = 9,325$ students from 135 different schools	Quality positive
Dufo <i>et al.</i> (2011)	Primary	Quantitative	2005–2006	Longitudinal	RCTs	121 schools. $n = 10,000$ students	Quality positive
Pierce <i>et al.</i> (2011)	Primary	Quantitative	2004–2006	Longitudinal	quasi-experimental	Year 1: $n = 161$ across 52 schools. Year 2: $n = 127$	Quality positive
Sorensen <i>et al.</i> (2017)	Secondary	Quantitative	2005–2006 to 2011–2012	Longitudinal	Non-experimental	$n =$ over 1.7 million 2,000 schools	Quality positive
Boaler and Foster (2021)	Primary	Quantitative qualitative	2005–2009 2013–2015	Longitudinal	quasi-experimental	Study 1:8 intervention districts and 25 comparison districts; Study 2: $n =$ over 11,000 students in 120 school districts	Quality negative
Burris <i>et al.</i> (2006)	Secondary	Quantitative	1995–2000	Longitudinal	Quasi-experimental	$n = 477$ students (pre-universal acceleration) and $n = 508$ post-universal acceleration	Quality negative
Deunk <i>et al.</i> (2018)	Secondary: meta-analysis	Quantitative	Late 1990s and early 2000s	Not specified for each study	Meta-analysis	$n = 5$ studies for mathematics	Quality negative
Gupta <i>et al.</i> (2023)	Primary	Quantitative	2015–2016	Longitudinal	Quasi-experimental	$n = 9170$ students across 19 schools from 23 counties	Quality negative

Source(s): Author's own work

Table 5. Study design, methodology and the impact of ability grouping on equity and inclusion

Author publication year	Data source	Type of data	Data recency	Time horizon	Study design	Sample size	Focus and force
Archer <i>et al.</i> (2018)	Secondary	Quantitative, qualitative	2015–2016	Cross-sectional	Non-experimental	Survey: $n = 12,935$ in 94 schools, Interviews $n = 33$ students	Equity negative
Hantas (2018)	Secondary	Quantitative	2012–2013	Cross-sectional	Non-experimental	$n = 9,610$ students	Equity negative
Von Hippel and Cañedo (2022)	Primary	Quantitative	2010–2011	Longitudinal	Non-experimental	fall: $n = 2,607$ students, and for spring $n = 1,355$ students	Equity negative
Gupta <i>et al.</i> (2023)	Primary	Quantitative	2015–2016	Longitudinal	Quasi-experimental	$n = 9,170$ students across 19 schools from 23 countries	Inclusion positive
Boliver and Capsada-Munsech (2021)	Secondary	Quantitative	2008 and 2012	Longitudinal	Non-experimental	$n = 8,876$ students	Inclusion negative
Campbell (2021)	Secondary	Quantitative	2008 and 2012	Longitudinal	Non-experimental	$n = 4,463$ students	Inclusion negative
Francome and Hewitt (2018)	Primary	Quantitative, qualitative	2014	Cross-sectional	Non-experimental	$n = 286$ year seven students in School M ($n = 129$ students) and School S ($n = 157$ students)	Inclusion negative
McDool (2019)	Secondary	Quantitative	2008 and 2012	Longitudinal	quasi-experimental	waves 4 (age 7): $n = 14,043$ students wave 5 (age 11 years): $n = 13,469$ students	Inclusion negative
Legette and Kurtz-Costes (2020)	Primary	Quantitative	2016–2017	Longitudinal	Non-experimental	$n = 322$ students from 4 schools	Inclusion mixed
Parker <i>et al.</i> (2021)	Secondary	Quantitative	2013–2015	Cross-sectional	Non-experimental	$n = 645,520$ from 22,894 schools	Inclusion mixed

Source(s): Author's own work

grouping on academic achievement, while four showed a positive effect. Most studies highlighted adverse effects on equity and inclusion, as indicated in Table 5. However, simply counting studies is insufficient, as their rigor and reliability vary based on their design and methodology. Therefore, the points system presented in Tables 6 and 7 is crucial for evaluating and comparing the impact of these studies more effectively.

Tables 6 and 7 present the findings from the force field analysis steps, which evaluate and quantify the strength of each study’s design and methodology. This analysis calculated each study’s total “force” (or strength) and aligned the direction of this force with the study’s conclusions. Each study was classified according to its relevance to the three primary themes: quality (Table 6), equity and inclusion (Table 7).

Table 6. Forces assessment for studies focusing on quality (measured by academic achievement)

Author publication year	Data source	Type of data	Data recency	Time horizon	Study design	Sample size	Total force	Focus and force
Collins and Gan (2013)	2	1.5	1	3	3.5	3	14	Quality positive
Duflo <i>et al.</i> (2011)	2.5	1.5	1	3	4	3	15	Quality positive
Pierce <i>et al.</i> (2011)	2.5	1.5	1	3	3.5	1	12.5	Quality positive
Sorensen <i>et al.</i> (2017)	2	1.5	1	3	2	3	12.5	Quality positive
Total quality positive							54	
Boaler and Foster (2021)	2.5	3	2.5	3	3.5	3	17.5	Quality negative
Gupta <i>et al.</i> (2023)	2.5	1.5	2.5	3	3.5	3	16	Quality negative
Burris <i>et al.</i> (2006)	2	1.5	1	3	3.5	3	14	Quality negative
Deunk <i>et al.</i> (2018)	2	1.5	1	1.5	4	1	11	Quality negative
Total quality negative							58.5	

Source(s): Author’s own work

Table 7. Forces assessment for studies focusing on equity and inclusion

Author publication year	Data source	Type of data	Data recency	Time horizon	Study design	Sample size	Total force	Focus and force
Archer <i>et al.</i> (2018)	2	3	2.5	1.5	2	3	14	Equity negative
Hartas (2018)	2	1.5	1	1.5	2	3	11	Equity negative
Von Hippel and Cañedo (2022)	2.5	1.5	1	3	2	3	13	Equity negative
Total equity negative							38	
Gupta <i>et al.</i> (2023)	2.5	1.5	2.5	3	3.5	3	16	Inclusion positive
Total inclusion positive							29	
Boliver and Capsada-Munsech (2021)	2	1.5	1	3	2	3	12.5	Inclusion negative
Campbell (2021)	2	1.5	1	3	2	3	12.5	Inclusion negative
Francome and Hewitt (2018)	2.5	3	2.5	1.5	2	3	14.5	Inclusion negative
McDool (2019)	2	1.5	1	3	3.5	3	14	Inclusion negative
Total inclusion negative							53.5	
Legette and Kurtz-Costes (2020)	2.5	1.5	2.5	3	2	1	12.5	Inclusion mixed
Parker <i>et al.</i> (2021)	2	1.5	2.5	1.5	2	3	12.5	Inclusion mixed
Total inclusion mixed							25	

Source(s): Author’s own work

Table 8 summarizes how ability grouping may function as a driving or restraining force in achieving inclusion, equity and quality. Studies that identified a positive impact of ability grouping were included in the total for the driving forces, while those indicating a negative impact were counted among the restraining forces.

The weighted average forces indicate the overall direction and strength of influence in each theme, accounting for all forces, including neutral and mixed ones. Including these in the weighted average calculation helps avoid overestimating positive or negative influences.

In the quality theme, the negative forces (58.5) slightly outweighed the positive forces (54), resulting in a weighted average of -0.04 . This result suggests a slight overall negative finding regarding the impact of ability grouping on academic achievement, indicating no consensus among the studies on whether ability grouping is beneficial or detrimental.

In the equity theme, there were no neutral or positive forces; the only influence stemmed from negative restraining forces (38). The resulting weighted average of -1 signifies a consensus among the studies that ability grouping is a restraining force in achieving equity.

In the inclusion category, both negative and mixed forces were observed. The negative restraining forces (53.5) significantly outweighed the positive driving ones (0), leading to a net negative force of -53.5 . When mixed forces (25) were included, the weighted average force was -0.682 , indicating that while some studies found ability grouping beneficial for some groups of students, most suggested that it acts as a restraining force, negatively affecting students' non-cognitive skills.

The total force aggregated forces across all categories. Negative restraining forces (150) were substantially higher than positive driving ones (54), resulting in a net negative force of -146.5 . Including mixed forces (25) yielded a weighted average force of -0.613 , suggesting that ability grouping has restraining properties for achieving SDG4. However, ability grouping was not universally detrimental; some studies indicated that homogeneous grouping can enhance academic achievement for specific student groups. Additionally, some scholars argue that equity in mathematics education should address the uniqueness of gifted and disadvantaged students (Leikin, 2011; Powell, 2015).

Discussion

This study reviewed the literature on ability grouping to understand better whether it acts as a restraining or driving force for achieving SDG4. Such analysis can aid schools and educational systems to challenge ineffective practices and make evidence-based decisions.

Some research suggests that ability grouping enhances academic achievement across all student levels by narrowing the range of abilities within classrooms (Collins and Gan, 2013; Duflo *et al.*, 2011). However, the equilibrium between driving and restraining forces did not universally support the positive impact of ability grouping. Boaler and Foster (2021) and

Table 8. Summary table: Achievement, equity and inclusion total forces

	Neutral and mixed driving forces	Positive driving force	Negative restraining force	Total forces	Positive – negative forces	Weighted average forces
Quality	0	54	58.5	112.5	-4.5	-0.04
Equity	0	0	38	48	-48	-1
Inclusion	25	0	53.5	78.5	-53.5	-0.682
Total force	25	54	150	239	-146.5	-0.613

Source(s): Author's own work

Burris *et al.* (2006) argued that ability grouping may restrict opportunities for lower-achieving students, exacerbating inequities and limiting their potential.

The analysis indicates that the impact of ability grouping on equity serves as a clear restraining force. Ability grouping acts as a barrier to achieving equitable outcomes in mathematics education, as evidenced by Archer *et al.* (2018), who highlighted that privileged students – especially those from higher socioeconomic backgrounds – were disproportionately placed in top-ability groups, while marginalized students from disadvantaged backgrounds were often relegated to lower groups. Similarly, Hartas (2018) found that teacher perceptions significantly influenced group placements, with biases against students from lower socioeconomic backgrounds and minority groups resulting in inequitable outcomes. Von Hippel and Cañedo (2022) further underscored the influence of social biases, noting that students from high socioeconomic backgrounds were more likely to be placed in higher groups than their performance alone suggested. These studies collectively demonstrated that ability grouping often reinforces existing inequities (Archer *et al.*, 2018; Hartas, 2018; Von Hippel and Cañedo, 2022), acting as a restraining force against achieving SDG4.

Regarding inclusion, the total negative forces outweighed the positive forces, resulting in a net restraining force. Although some studies (Gupta *et al.*, 2023) indicated potential positive effects of ability grouping on some students' non-cognitive skills, others revealed that students placed in lower-ability groups often suffered negative impacts on their motivation and sense of belonging. Indeed, ability grouping, particularly at the primary level, was associated with declining students' enjoyment of mathematics, especially for those in lower groups (Boliver and Capsada-Munsech, 2021). Francome and Hewitt (2018) reported that homogeneous ability grouping limited opportunities for exploration and collaboration, fostering more procedural learning rather than deep engagement.

Regarding self-concept, the findings are not as clear. Campbell (2021) found that girls in the lowest groups were likelier to develop negative mathematics self-concept later. Legette and Kurtz-Costes (2020) and Parker *et al.* (2021) had mixed results. Legette and Kurtz-Costes (2020) found that students in honors mathematics classes reported higher math self-concept than their peers in regular classes, suggesting that ability grouping exacerbates differences in self-concept. However, lower-achieving students in regular classes had a lower self-concept than their peers, indicating that ability grouping may have positive and negative impacts depending on the group.

In contrast, Parker *et al.* (2021) found that higher-achieving students experienced a diminished self-concept, while lower-achieving students developed an inflated self-concept because of the BFLPE. This study was not rated as negative, even if this inflated self-concept does not necessarily lead to greater academic mobility. It was rated mixed in the force field analysis because the impact on self-concept was the factor being analyzed.

The partial force field analysis findings indicate that ability grouping restrains SDG4 achievement. However, this does not imply that ability grouping is universally detrimental. Some scholars argue that equity in mathematics education should account for both gifted and disadvantaged students' unique needs (Leikin, 2011; Powell, 2015). These findings emphasize the need to explore alternative practices, such as differentiated instruction or flexible grouping (Boaler and Foster, 2021; Burris *et al.*, 2006), to ensure equitable and inclusive access to mathematics and STEM education for all students.

Limitations

Reliance on English-language publications and the SCOPUS and ERIC databases may have introduced selection bias by excluding relevant studies published in other languages or indexed in different databases. Additionally, using specific keywords ("ability AND

grouping AND math”) may have further limited the scope, as the search predominantly returned studies on between- and within-class grouping. Given the variation in terminology across educational systems – with terms such as tracking, streaming, setting and sorting often used interchangeably – other forms of grouping may have been unintentionally excluded. Studies explicitly addressing between-school tracking were underrepresented. While [Parker et al. \(2021\)](#) included between-school grouping, the data in their study did not allow for a clear distinction between within- and between-school tracking. The geographical focus is also skewed toward Western settings, limiting generalizability. Methodologically, the predominance of non-experimental studies prevents causal conclusions. Additionally, while qualitative research on ability grouping exists, only three were retrieved in this review, limiting insights into students’ lived experiences.

While a points system was used to assess study rigor, some subjectivity remains in categorization, and different weighting schemes could yield different results. This approach was necessary because simply counting studies with positive or negative effects would not account for variations in research quality.

Conclusion

This study reviewed the literature on ability grouping in mathematics education and applied force field analysis to assess its role as a driving or restraining force in achieving SDG4 goals of inclusion, equity and quality. To fully align with SDG4’s holistic vision, schools must ensure that student grouping practices positively support each of these components. Findings suggest ability grouping restrains SDG4 progress. While its impact on academic achievement remains debated, concerns persist over placement bias and negative effects on students’ non-cognitive outcomes.

As [Sukhnandan and Lee \(1998\)](#) note, mathematics is often seen as well-suited for ability grouping because its structured, sequential nature supports differentiated pacing for varying skill levels. However, continued research is needed to determine whether contemporary educational goals – emphasizing critical thinking, problem-solving and depth of understanding ([NCTM, 2023b](#)) – justify continuing with traditional ability grouping. Notably, this study indicates that mathematics placement decisions can be biased toward perceived versus students’ actual ability.

Moreover, some scholars argue that equity in mathematics education should consider the unique needs of both gifted and disadvantaged students ([Leikin, 2011](#); [Powell, 2015](#)). Therefore, examining how schools define inclusion, equity and quality in mathematics education is essential. Establishing a shared vision for these components is crucial for advancing SDG4 and enabling educators to make informed decisions about implementing ability grouping in ways that enhance both the breadth and depth of students’ mathematical learning.

References

- Allen, K.A., Greenwood, C.J., Berger, E., Patlamazoglou, L., Reupert, A. and Wurf, G. (2024), “Adolescent school belonging and mental health outcomes in young adulthood: findings from a multi-wave prospective cohort study”, *School Mental Health*, Vol. 16 No. 1, pp. 149-160, doi: [10.1007/s12310-023-09626-6](https://doi.org/10.1007/s12310-023-09626-6).
- Archer, L., Francis, B., Miller, S., Taylor, B., Tereshchenko, A., Mazonod, A.V., Pepper, D. and Travers, M.C. (2018), “The symbolic violence of setting: a Bourdieusian analysis of mixed methods data on secondary students’ views about setting”, *British Educational Research Journal*, Vol. 44 No. 1, pp. 119-140, doi: [10.1002/berj.3321](https://doi.org/10.1002/berj.3321).

- Boaler, J. (2015), "Fluency without fear: research evidence on the best ways to learn math facts", *Stanford News*, January 29, available at: <https://news.stanford.edu/stories/2015/01/math-learning-boaler-012915>
- Boaler, J. (2020), "Ability grouping in mathematics classrooms", in Lerman, S. (Ed.), *Encyclopedia of Mathematics Education*, Springer, Cham, doi: [10.1007/978-3-030-15789-0_145](https://doi.org/10.1007/978-3-030-15789-0_145).
- Boaler, J. (2024), *Math-Ish: Finding Creativity, Diversity, and Meaning in Mathematics*, HarperOne, San Francisco, CA.
- Boaler, J. and Foster, D. (2021), "Raising expectations and achievement: the impact of two wide-scale de-tracking mathematics reforms", *Stanford University and Silicon Valley*.
- Boeren, E. (2019), "Understanding sustainable development goal (SDG) 4 on 'quality education' from micro, meso, and macro perspectives", *International Review of Education*, Vol. 65 No. 2, pp. 277-294, doi: [10.1007/s11159-019-09772-7](https://doi.org/10.1007/s11159-019-09772-7).
- Boliver, V. and Capsada-Munsech, Q. (2021), "Does ability grouping affect UK primary school pupils' enjoyment of Maths and English?", *Research in Social Stratification and Mobility*, Vol. 76, p. 100629, doi: [10.1016/j.rssm.2021.100629](https://doi.org/10.1016/j.rssm.2021.100629).
- Borghans, L., Duckworth, A., Heckman, J. and ter Weel, B. (2008), "The economics and psychology of personality traits", *Journal of Human Resources*, Vol. 43 No. 4, pp. 972-1059.
- Burris, C.C., Heubert, J.P. and Levin, H.M. (2006), "Accelerating mathematics achievement using heterogeneous grouping", *American Educational Research Journal*, Vol. 43 No. 1, pp. 137-154, doi: [10.3102/00028312043001105](https://doi.org/10.3102/00028312043001105).
- Cameron, E. and Green, M. (2024), *Making Sense of Change Management: A Complete Guide to the Models, Tools and Techniques of Organizational Change* (6th ed.), Kogan Page Publishers, London.
- Campbell, T. (2021), "In-class' ability'-grouping, teacher judgments and children's mathematics self-concept: evidence from primary-aged girls and boys in the UK millennium cohort study", *Cambridge Journal of Education*, Vol. 51 No. 5, pp. 563-587, doi: [10.1080/0305764X.2021.1877619](https://doi.org/10.1080/0305764X.2021.1877619).
- Chiu, M., Chow, B., McBride, C. and Mol, S. (2016), "Students' sense of belonging at school in 41 countries: cross-cultural variability", *Journal of Cross-Cultural Psychology*, Vol. 47, pp. 10-1177, doi: [10.1177/0022022115617031](https://doi.org/10.1177/0022022115617031).
- Chmielewski, A.K. (2014), "An international comparison of achievement inequality in within- and between-school tracking systems", *American Journal of Education*, Vol. 120 No. 3, pp. 293-324, doi: [10.1086/675529](https://doi.org/10.1086/675529).
- Collins, C.C. and Gan, L. (2013), "Does sorting students improve scores? An analysis of class composition", Cambridge, MA: National Bureau of Economic Research, available at: www.nber.org/papers/w18848 (accessed 27 December 2024).
- Creswell, J.W. (2021), *A Concise Introduction to Mixed Methods Research*, Sage Publications, Thousand Oaks, CA.
- Creswell, J.W. and Creswell, J.D. (2018), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Sage Publications, Los Angeles, CA.
- Dahlberg, O. (2021), "Att räkna eller räknas bort: om matematik, behörighet och ingenjörer [to count or Be counted out: on mathematics, eligibility, and engineers]", *Sveriges Ingenjörer*.
- Deunk, M.I., Smale-Jacobse, A.E., De Boer, H., Doolaard, S. and Bosker, R.J. (2018), "Effective differentiation practices: a systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education", *Educational Research Review*, Vol. 24, pp. 31-54, doi: [10.1016/j.edurev.2018.02.002](https://doi.org/10.1016/j.edurev.2018.02.002).
- Douglas, D. and Attewell, P. (2017), "School mathematics as gatekeeper", *The Sociological Quarterly*, Vol. 58 No. 4, pp. 648-669, doi: [10.1080/00380253.2017.1354733](https://doi.org/10.1080/00380253.2017.1354733).

- Duflo, E., Dupas, P. and Kremer, M. (2011), "Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya", *American Economic Review*, Vol. 101 No. 5, pp. 1739-1774, doi: [10.1257/aer.101.5.1739](https://doi.org/10.1257/aer.101.5.1739).
- Dweck, C.S., Walton, G.M. and Cohen, G.L. (2014), "Academic tenacity: mindsets and skills that promote long-term learning", Bill and Melinda Gates Foundation.
- Francome, T. and Hewitt, D. (2018), "My math lessons are all about learning from your mistakes: How mixed-attainment mathematics grouping affects the way students experience mathematics", *Educational Review*, Vol. 72 No. 4, pp. 475-494, doi: [10.1080/00131911.2018.1513908](https://doi.org/10.1080/00131911.2018.1513908).
- Gamoran, A. and Hallinan, M.T. (1995), "Tracking students for instruction", in Hallinan, M.T. (Ed.), *Restructuring Schools*, Springer, Boston, MA, doi: [10.1007/978-1-4899-1094-3_7](https://doi.org/10.1007/978-1-4899-1094-3_7).
- Goodenow, C. and Grady, K.E. (1993), "The relationship of school belonging and friends' values to academic motivation among urban adolescent students", *Journal of Experimental Education*, Vol. 62 No. 1, pp. 60-71, doi: [10.1080/00220973.1993.9943831](https://doi.org/10.1080/00220973.1993.9943831).
- Gupta, S., Liu, C., Li, S., Chang, F. and Shi, Y. (2023), "Association between ability tracking and student's academic and non-academic outcomes: empirical evidence from junior high schools in rural China", *International Journal of Educational Development*, Vol. 103, p. 102927, doi: [10.1016/j.ijedudev.2023.102927](https://doi.org/10.1016/j.ijedudev.2023.102927).
- Hanushek, E.A. and Woessmann, L. (2006), "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries", *The Economic Journal*, Vol. 116 No. 510, pp. C63-C76, doi: [10.1111/j.1468-0297.2006.01076.x](https://doi.org/10.1111/j.1468-0297.2006.01076.x).
- Hartas, D. (2018), "Setting for English and Maths: 11-year-olds' characteristics and teacher perceptions of school attitudes", *Research Papers in Education*, Vol. 33 No. 3, pp. 393-410, doi: [10.1080/02671522.2017.1329338](https://doi.org/10.1080/02671522.2017.1329338).
- Institute of Education Sciences (IES) (2025), "ESSA evidence tiers: video handout", available at: <https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf> (accessed 27 December 2024).
- Jamali, S.M., Nader, A.E. and Jamali, F. (2022), "The role of STEM education in improving the quality of education: a bibliometric study", *International Journal of Technology and Design Education*, Vol. 33 No. 3, pp. 819-840, doi: [10.1007/s10798-022-09762-1](https://doi.org/10.1007/s10798-022-09762-1).
- Legette, K.B. and Kurtz-Costes, B. (2020), "Math track placement and reflected classroom appraisals are related to changes in early adolescents' math self-concept", *Educational Psychology*, Vol. 41 No. 5, pp. 602-617, doi: [10.1080/01443410.2020.1760212](https://doi.org/10.1080/01443410.2020.1760212).
- Leikin, R. (2011), "The education of mathematically gifted students: some complexities and questions", *The Mathematics Enthusiast*, Vol. 8 Nos 1/2, pp. 167-188, doi: [10.54870/1551-3440.1211](https://doi.org/10.54870/1551-3440.1211).
- Levinson, M., Geron, T. and Brighthouse, H. (2022), "Conceptions of educational equity", *AERA Open*, Vol. 8 No. 1, pp. 1-12, doi: [10.1177/23328584221121344](https://doi.org/10.1177/23328584221121344).
- Lewin, K. (1951), *Field Theory in Social Science*, Harper and Brothers, New York, NY.
- McDool, E. (2019), "Ability grouping and children's non-cognitive outcomes", *Applied Economics*, Vol. 52 No. 28, pp. 3035-3054, doi: [10.1080/00036846.2019.1705239](https://doi.org/10.1080/00036846.2019.1705239).
- Marsh, H.W., Kuyper, H., Morin, A.J.S., Parker, P.D. and Seaton, M. (2014), "Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications", *Learning and Instruction*, Vol. 33, pp. 50-66, doi: [10.1016/j.learninstruc.2014.04.002](https://doi.org/10.1016/j.learninstruc.2014.04.002).
- National Science Board (2024), "Accelerating STEM talent: an urgent priority for U.S. competitiveness (NSB-2024-4)", available at: www.nsf.gov/nsb/publications/2024/2024_policy_brief.pdf (accessed 27 December 2024).
- NCTM (2018), "Building STEM education on a sound mathematical foundation", available at: www.nctm.org/Standards-and-Positions/Position-Statements/Building-STEM-Education-on-a-Sound-Mathematical-Foundation/ (accessed 27 December 2024).

- NCTM (2023a), “Ability labels: Disrupting ‘high’, ‘medium’, and ‘low’ in mathematics education”, available at: www.nctm.org/Standards-and-Positions/Position-Statements/Ability-Labels_-_Disrupting-%E2%80%9CHigh,%E2%80%9D-%E2%80%9CMedium,%E2%80%9D-and-%E2%80%9CLow%E2%80%9D-in-Mathematics-Education/ (accessed 27 December 2024).
- NCTM (2023b), “The effective and appropriate use of large-scale assessments in mathematics education to guide systemic improvement and equitable student learning”, available at: www.nctm.org/Standards-and-Positions/Position-Statements/The-Effective-and-Appropriate-Use-of-Large-Scale-Assessments-in-Mathematics-Education-to-Guide-Systemic-Improvement-and-Equitable-Student-Learning/ (accessed 21 December 2024).
- NSF (2020), “STEM education for the future: a visioning report”, available at: www.nsf.gov/edu/Materials/STEM%20Education%20for%20the%20Future%20-%202020%20Visioning%20Report.pdf (accessed 27 December 2024).
- Parker, P., Dicke, T., Guo, J., Basarkod, G. and Marsh, H. (2021), “Ability stratification predicts the size of the big-fish-little-pond effect”, *Educational Researcher*, Vol. 50 No. 6, pp. 334-344, doi: [10.3102/0013189X20986176](https://doi.org/10.3102/0013189X20986176).
- Pierce, R.L., Cassady, J.C., Adams, C.M., Neumeister, K.L.S., Dixon, F.A. and Cross, T.L. (2011), “The effects of clustering and curriculum on the development of gifted learners’ math achievement”, *Journal for the Education of the Gifted*, Vol. 34 No. 4, pp. 569-594, doi: [10.1177/016235321103400403](https://doi.org/10.1177/016235321103400403).
- Powell, S.R. (2015), “Connecting evidence-based practice with implementation opportunities in special education mathematics preparation”, *Intervention in School and Clinic*, Vol. 51 No. 2, pp. 90-96, doi: [10.1177/1053451215579269](https://doi.org/10.1177/1053451215579269).
- Redmond-Sanogo, A., Angle, J. and Davis, E. (2016), “Kinks in the STEM pipeline: tracking STEM graduation rates using science and mathematics performance”, *School Science and Mathematics*, Vol. 116 No. 7, pp. 378-388, doi: [10.1111/ssm.12195](https://doi.org/10.1111/ssm.12195).
- Seeley, C.L. (2004), “Hard arithmetic is not deep mathematics”, *National Council of Teachers of Mathematics*, available at: www.nctm.org/uploadedFiles/News_and_Calendar/Messages_from_the_President/Archive/Cathy_Seeley/2004_10_hardarithmetic.pdf (accessed 15 March 2025).
- Slavin, R.E. (1987), “Ability grouping and student achievement in elementary schools: a best-evidence synthesis”, *Review of Educational Research*, Vol. 57 No. 3, pp. 293-336, doi: [10.2307/1170460](https://doi.org/10.2307/1170460).
- Snyder, K.M. (2023), “Expanding how we think about quality in education”, in Snyder, K.J. and Snyder, K.M. (Eds), *Systems Thinking for Sustainable Schooling: A Mindshift for Educators to Lead and Achieve Quality Schools*, Rowman and Littlefield Publishing Group, Lanham, MD, USA, pp. 25-42.
- Snyder, K.J. and Anderson, R.H. (1986), *Managing Productive Schools: Toward an Ecology*, Academic Press, Orlando, FL, p. 479.
- Sorensen, L.C., Cook, P.J. and Dodge, K.A. (2017), “From parents to peers: trajectories in sources of academic influence grades 4 to 8”, *Educational Evaluation and Policy Analysis*, Vol. 39 No. 4, pp. 697-711, doi: [10.3102/0162373717708335](https://doi.org/10.3102/0162373717708335).
- Steenbergen-Hu, S., Makel, M.C. and 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*, Vol. 86 No. 4, pp. 849-899, doi: [10.3102/0034654316675417](https://doi.org/10.3102/0034654316675417).
- Sukhmandan, L. and Lee, B. (1998), “Streaming, setting, and grouping by ability: a review of the literature”, National Foundation for Educational Research (NFER).
- UNESCO (2015), “SDG4-education 2030, Incheon declaration (ID) and framework for action. For the implementation of sustainable development goal 4”, *Ensure Inclusive and Equitable Quality Education and Promote Lifelong Learning Opportunities for All* (ed. 2016/WS/28).

UNESCO (2017), “A guide for ensuring inclusion and equity in education”, United Nations Educational, Scientific and Cultural Organization, doi: [10.54675/MHHZ2237](https://doi.org/10.54675/MHHZ2237).

UNESCO (2020), “Global education monitoring report 2020: inclusion and education: all means all”, UNESCO, doi: [10.54676/JJNK6989](https://doi.org/10.54676/JJNK6989).

Von Hippel, P.T. and Cañedo, A.P. (2022), “Is kindergarten ability group placement biased? New data, new methods, new answers”, *American Educational Research Journal*, Vol. 59 No. 4, pp. 820-857, doi: [10.3102/00028312211061410](https://doi.org/10.3102/00028312211061410).

Further reading

OECD (2016), *PISA 2015 Results: Policies and Practices for Successful Schools*, Vol. 2, OECD Publishing, Paris, doi: [10.1787/9789264267510-en](https://doi.org/10.1787/9789264267510-en).

About the author

Telma Largent is a Doctoral Student at Mid Sweden University. Born in Brazil, she currently lives in the USA while completing her studies remotely. Her research investigates the forces that contribute to resistance to change in mathematics education. The purpose of her study is to identify and assess driving and restraining forces that can inform the planning of effective strategies promoting access to equitable and inclusive quality mathematics education for all. Telma Largent can be contacted at: telma.largent@miun.se