

The Effect of Level-Marked Tasks on Students' Mathematics Self-Efficacy and Persistence—Investigating a Tiered Activity

Maria Herset¹, Annette H. Bjerke² & Mohamed El Ghami¹

¹ Faculty of Education and Arts, Nord University, Norway

² Faculty of Education and International Studies, OsloMet, Norway

Correspondence: Mohamed El Ghami, Faculty of Education and Arts, Nord University, 8700 Nesna, Norway. E-mail: mohamed.el-ghami@nord.no

Received: February 13, 2024; Accepted: February 28, 2024; Published: March 4, 2024

Abstract

In school mathematics, tiering tasks according to students' readiness levels is a common approach to adapting content to address students' varying abilities. To distinguish between different tiers, the tasks are commonly marked by their level of difficulty, as in so called level-marked tasks – a strategy used in differentiated instruction to help students gain mastery and improve performance. However, less is known about whether and to what extent the marking of the tasks works as intended. Regarded as a significant predictor of students' performance in mathematics, the level marking of mathematics tasks and their effects on students' self-efficacy and persistence were investigated. The online survey part of an experimental design was used to collect data from 436 lower secondary students in Norway. Independent sample t-tests and Mann–Whitney U tests revealed that, when students across groups were given the same mathematics tasks with and without level markings, tasks marked as difficult reduced both students' mathematics self-efficacy and their persistence, even for students with high baseline self-efficacy. Although more research is needed to fully understand how the level markings of mathematics tasks affect students' self-efficacy and persistence, the reported results have implications for both mathematics teachers and textbook authors.

Keywords: self-efficacy, persistence, differentiated instruction, mathematics education, level of difficulty, tiered activity

1. Introduction

Self-efficacy, initially defined by Bandura (1997) as “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (p. 3), is very advantageous to possess (Bandura, 1993). An ever-growing body of research has reported the outcomes associated with possessing self-efficacy, such as performance and persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Óturai et al., 2023; Pajares, 1996; Pajares & Miller, 1995; Schunk, 1991; Sexton & Tuckman, 1991; Skaalvik et al., 2015; Zakariya, 2021). In line with Sasidharan and Kareem (2023), we argue that in mathematics—a school subject that many students tend to perceive as difficult (Moyer et al., 2018)—high self-efficacy and persistence are of the utmost importance. If students' self-efficacy is high, they might exhibit a higher degree of participation in mathematics activities, which can increase their well-being (Sasidharan & Kareem, 2023). Therefore, we assert that it is important to provide students with opportunities to develop self-efficacy, and hence persistence, in mathematics.

Given that the experience of mastery when studying mathematics promotes self-efficacy in individuals (Bandura, 1997; Butz & Usher, 2015; Joët et al., 2011; Stevens et al., 2006; Usher & Pajares, 2009; Zientek et al., 2019), which mediates persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Multon et al., 1991; Lyman et al., 1984; Schunk, 1991; Sexton & Tuckman, 1991; Skaalvik et al., 2015), an important part of the puzzle is for teachers to facilitate mastery. In this regard, differentiated instruction, understood as a flexible way to tailor teaching to meet individual needs (Tomlinson, 2014), is key in mathematics education (Bal, 2016; Luster, 2008; Scott, 2012; Suarez, 2007). Differentiated instruction has been found to be particularly important in lower secondary school (Tomlinson, 2005), where it has been argued that balancing the challenge of a task with a student's ability level can be decisive in a particular student's progress (Nakamura & Csikszentmihalyi, 2002). This strategy is called *tiering according to readiness* (Pierce & Adams, 2005). To distinguish between different tiers (e.g., below grade level, at grade level, and above grade level), mathematics tasks can be marked or grouped

according to the tasks' level of difficulty (Grave & Pepin, 2015). Such tasks are called level-marked tasks, employed in tiered activities, and are often found in mathematics textbooks (Auliya & Widjajanti, 2023; Brändström, 2005; Grave & Pepin, 2015), and implemented in classroom practices (Eriksen et al., 2022; Grave & Pepin, 2015; Suarez, 2007).

The focus of this paper is *level marking of mathematics tasks*. To be clear, we are not investigating how task difficulty levels affect students' self-efficacy and persistence, but rather examining how different labels themselves (easy, medium, and difficult) affect students' self-efficacy and persistence.

To our knowledge, even though level marking of mathematics tasks is a widespread approach, little research has reported on the consequences of marking tasks with different levels of difficulty. Therefore, in this paper, we report on a study from Norway investigating how marking tasks as easy, medium, and difficult affects students' self-efficacy and persistence in mathematics. We assert that this study is important for mathematics teachers to understand students' self-efficacy and persistence when level-marked tasks are used in tiered activities. In addition, the results may be useful for policymakers, textbook authors, and teachers when they have to make suggestions about whether we should continue, or implement, the use of level-marked tasks as a strategy to help students gain mastery and improve their performance.

1.1 Self-Efficacy and Level Marking of Mathematics Tasks

Bandura (1997) distinguished between perceived self-efficacy and outcome expectations. While outcome expectations are a person's judgment of the likely consequences that a particular performance will produce (Bandura, 1997, p. 21), in this paper, we focus on students' perceived mathematics self-efficacy, which is understood as "a situational or problem-specific assessment of an individual's confidence in her or his ability to successfully perform or accomplish a particular task or problem" (Hackett & Betz, 1989, p. 262).

Self-efficacy is a multifaceted phenomenon that varies in strength, level, and generality (Bandura, 1997). In this paper, we specifically measure students' strength of self-efficacy, which refers to how strong a person's belief is in completing a given task (Bandura, 1997). According to Street et al. (2017), students' self-efficacy strength is affected by how they assess a specific task (generality) and how difficult they perceive the task to be (level). This is supported by other researchers, who found that students' self-efficacy is affected by a task's difficulty level (Chen & Zimmerman, 2007) and that students who find a given task difficult report lower self-efficacy compared to students who find the task easy (Collins, 1984; Liu et al., 2020). However, how students perceive a task's difficulty level can differ between students, and such perceptions do not necessarily reflect an "objective" level of difficulty defined by a teacher (Street et al., 2017). Liu et al. (2020) found that even if students reported higher self-efficacy when they were given easy tasks (marked as easy) compared to when they were given difficult tasks (marked as difficult), there was a "difficulty-efficacy mismatch" (p. 419), as their performance was lower for easy tasks compared to that for difficult tasks. From this, the researchers concluded that students' self-efficacy can be affected by perceived task difficulty and not by assessments of tasks in relation to one's own capabilities. Taken together, from this, we can infer that when given information about a task's difficulty level (as in the level marking of mathematics tasks), this may affect how difficult students perceive a task to be (level of self-efficacy), which again may affect their strength of self-efficacy. However, to the best of our knowledge, this phenomenon has not yet been researched.

1.2 Persistence and Level Marking of Mathematics Tasks

Building on a recent literature review on the issue, DiNapoli (2023) suggested that persistence should be defined as the "voluntary continuation of a goal-directed action in spite of obstacles, difficulties, or discouragement," a definition originally set forth by Peterson and Seligman (2004, p. 229). Persistence is linked to students' short-term goals (e.g., solving a difficult mathematics task) and not long-term goals (e.g., getting a master's degree), similar to the idea of grit (DiNapoli, 2023). According to DiNapoli (2023), focusing on persistence when studying students' habits empirically will provide information about students' engagement time. In particular, in experimental studies like ours, DiNapoli (2023) found that researchers often operationalize persistence as "the amount of time spent on working on a difficult task" (p. 20).

Considering this operationalization of persistence, a task's degree of difficulty is decisive in how best to measure persistence, especially considering how "more competent individuals may find solutions more quickly than less competent individuals" (Shen et al., 2021, p. 42). Multon et al. (1991) and DiNapoli (2023) suggested that persistence should be measured during challenging circumstances. This is in line with Montague and Applegate (2000), who found that when a task is easy, students who struggle with mathematics spend more time trying to solve the task compared to high achievers, and when a task is hard, the first group of students tends to give up, while high achievers persist longer. Similarly, in a study conducted by Spielberg and Azaria (2021), students'

persistence was examined by revealing the true difficulty of some riddle tasks. The results revealed no statistically significant difference between students who received easy tasks marked as easy compared to students who received the same tasks without information about the level of difficulty. In addition, when difficult tasks were marked according to the level of difficulty, students' time spent coming up with a solution was significantly higher compared to when students were given the same tasks without a difficulty level marking. Interestingly, the students' solution ratio was not influenced by revealing the tasks' true difficulty level. From this, we inferred that when a task is solvable, as in Spielberg and Azaria (2021), it is difficult to conclude whether a longer amount of time spent solving a task means that students are more persistent or less efficient. Hence, to avoid this confounder, the task needs to be impossible to solve. Therefore, a common approach is to measure persistence by recording the time students spend attempting to solve an impossible task (e.g., Brown & Inouye, 1978; Collins, 1984; Jacobs et al., 1984; Shen et al., 2016). As suggested by Stein and Burchartz (2006), an impossible task is one whose only solution is "there is no solution."

1.3 Persistence and Self-Efficacy

A large body of research has examined the relationship between self-efficacy and persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Schunk, 1991; Sexton & Tuckman, 1991; Skaalvik et al., 2015), but few have investigated this relationship in the context of mathematics tasks using an experimental design. Three rare examples come from Sexton and Tuckman (1991), Collins (1984), and Jacobs et al. (1984). Sexton and Tuckman (1991) measured self-efficacy and persistence when completing a series of mathematics tasks. At the beginning of the survey, generalized self-efficacy and previous performance were the best predictors of persistence, but at the end of the survey, specific self-efficacy and recent experiences completing the survey were the best predictors of persistence (Sexton & Tuckman, 1991). In other words, experience completing the survey changed the way students behaved, and their new behavioral patterns served as a source that influenced their choice to persist in the future. In the study conducted by Collins (1984), students were divided into three achievement groups (low, average, and high achievers) and divided again into two subgroups according to their self-efficacy (high or low). All students were given two impossible tasks. Collins (1984) found that for low- and average-achieving groups, students with high self-efficacy spent more time trying to solve the tasks compared to students with low self-efficacy, while for students in the high-achievement group, those with low self-efficacy spent more time solving the tasks compared to students with high self-efficacy. Jacobs et al. (1984) conducted an experimental study and found a strong effect of self-efficacy on students' persistence when they were asked to solve an impossible task that required them to make a tracing that combined a set of lines without lifting the pen from the paper. The students' self-efficacy was manipulated, with one group being told that the task was easy and another that the task was difficult. The students' persistence was measured by the amount of time they spent working on the task. Students who were told that the task was difficult reported lower self-efficacy than those who were told the task was easy. In addition, the results indicated that students who were told that they would do well on the task persisted significantly longer than those who were told the opposite. The last study in this list was the one we found to be most similar to ours, however, Jacobs et al. (1984) did not include a control group where participants were not provided any information about the difficulty of the task. None of the three examples discussed included level markings in their designs.

1.4 The Present Study

In the present study, we investigated how the level marking of mathematics tasks according to different labels affect students' self-efficacy and persistence. We examined different labels for the same task because previous research found that the level markings of mathematics tasks in textbooks are not always correct (Brändström, 2005; Singh, 2017). In addition, it can be problematic to mark tasks according to levels of difficulty; Krauthausen (2018) claimed that the level of difficulty is a subjective evaluation. From this, we learned that the level marking of a task might not always be perceived as correct.

In the beginning of the survey, the task had no level marking, and students' self-efficacy, related to this task, refers to as students' "«baseline self-efficacy»". Considering students' baseline self-efficacy (low or high), we also investigated how the level marking affected student's self-efficacy and persistence. Hence, we raised the following four research questions (RQs):

- RQ₁: How does the level marking of a mathematics task that is impossible to solve affect students' self-efficacy?
- RQ₂: Considering students' baseline self-efficacy, how does the level marking of a mathematics task that is impossible to solve affect students' self-efficacy?
- RQ₃: How does the level marking of a mathematics that is impossible to solve affect students' persistence?

RQ4: Considering students' baseline self-efficacy, how does the level marking of a mathematics task that is impossible to solve affect students' persistence?

Given that previous research found a negative effect of "difficult" marking of tasks (which were considered easy) on students' self-efficacy (Herset et al., 2023) and a positive effect on students' self-efficacy when they were told the task was easy (Jacobs et al., 1984), we put forward the following two directional hypotheses in accordance with RQ1:

H_{1A}: Students' self-efficacy is significantly reduced when they are faced with the same impossible task twice, first without a level marking and next with "difficult" marking.

H_{1B}: Students' self-efficacy is significantly increased when they are faced with the same impossible task twice, first without a level marking and next with an "easy" or "medium" marking.

Since we were unable to find previous research on the relationship between task level marking and baseline self-efficacy on students' self-efficacy, and since students' perceived task difficulty level might differ from the task's actual difficulty level, which is important when students form their efficacy beliefs (Street et al., 2017, 2022), we put forward the following two non-directional hypotheses in accordance with RQ2:

H_{2A}: For students with low baseline self-efficacy, the level marking of an impossible task affects their self-efficacy.

H_{2B}: For students with high baseline self-efficacy, the level marking of an impossible task affects their self-efficacy.

To study students' behavioral persistence, we used an impossible mathematics task. This extends previous research on the effect of level-marked tasks on students' self-efficacy and persistence. We expected the effect of a "difficult" marking on time spent on an impossible task to be negative (Herset & El Ghami, 2022; Jacobs et al., 1984) and the effect of an "easy" marking on persistence to be positive (Jacobs et al., 1984). However, there is also support for the opposite (Spielberg & Azaria, 2021). Since the literature is not clear, in accordance with RQ3, we put forward a non-directional hypothesis:

H₃: The marking of an impossible mathematics task affects students' time spent on the task.

Previous research found that students' perceptions of a task's difficulty affect their self-efficacy (Street et al., 2017), and given the positive relationship between self-efficacy and persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Schunk, 1991; Skaalvik et al., 2015), we expected there to be a relationship between students' level of self-efficacy and the level marking of mathematics task on students' persistence. Therefore, we proposed two hypotheses in accordance with RQ4:

H_{4A}: For students with low baseline self-efficacy, the level marking of an impossible task affects their time spent on the task.

H_{4B}: For students with high baseline self-efficacy, the level marking of an impossible task affects their time spent on the task.

2. Materials and Methods

In this paper, we draw on a subset of data collected from a larger project that aimed to investigate if and how the level marking of mathematics tasks affects students' self-efficacy, persistence, performance, and choice of tasks. Acknowledging that "more competent individuals may find solutions more quickly than less competent individuals" (Shen et al., 2016, p. 42), a survey was created by the first author, inspired by the work of other researchers (Brown & Inouye, 1978; Jacobs et al., 1984; Miele et al., 2022; Shen et al., 2016). The survey included one impossible task placed at the end of the survey, and other tasks measuring factors such as self-efficacy. The purpose of the impossible task was to measure persistence. As in prior research (Brown & Inouye, 1978; Jacobs et al., 1984; Miele et al., 2022; Shen et al., 2016), we measured persistence by quantifying the time a student spent working on a given task. The task reads as follows:

Emil and Sandra went to the shop to buy some fruit. Emil bought two apples and one pear. He paid NOK 30. Sandra bought four apples and two pears. She paid NOK 40. How much does one apple cost?

There are, of course, ethical considerations in giving students an impossible task. We agreed, as did Borge (2003), that the utility of the research outweighed any discomfort or risk for the informants. Because solving systems of linear equations with two unknowns where the system has no solution (which is what this task requires of the students) is included in the students' curriculum, we decided that there was no risk of giving them such a task.

However, if any of the students felt any discomfort after encountering this task, their teachers were informed and ready to discuss the task with them after data collection.

2.1 Data Collection

To ensure we could make generalizations, we used probability and cluster sampling, the latter being justified by the fact that the population of Norway is large and widely dispersed. To avoid a close cluster sample (Cohen et al., 2018), we recruited schools randomly across Norway. We selected 23 public schools and tested all students aged 13–15 years in the selected schools. Unfortunately, because of COVID-19 and school lockdowns, we needed to replace some of the randomly chosen schools with schools closer to the first author’s university.

The schools were contacted by the first author, who called each of the chosen schools’ principals. If the principals were willing to encourage the school’s mathematics teachers to participate, the teachers also needed to agree. We ensured that no students would feel obligated to participate. To ensure that the data collection was done in the same way across all chosen schools, each mathematics teacher was given data collection instructions. For example, participants had to conduct the survey during a lesson (and not outside school) to make sure no one collaborated or used calculators. Before the data were collected, the Norwegian Social Science Data Service (previously NSD, now Sikt) approved the project.

2.2 Participants

In total, 436 students in grades 8 and 9 (i.e., ages 13–15 years) responded to the online survey. Some of the data were found to be incomplete or monotonous, indicating that participants had skipped items, so 82 responses were deleted. In addition, 14 response strings were detected as outliers because the participants had spent an excessive amount of time on the survey, and 9 were detected as outliers because the differences in students’ self-efficacy were considered extreme. The final sample included 331 students, with an equal distribution of gender, coming from 23 schools from all regions in Norway (4% from Southern Norway, 9% from Western Norway, 30% from Eastern Norway, 10% from Mid-Norway, and 47% from Northern Norway).

2.3 Experimental Design

The survey used in this study was designed for a larger research project and included 10 solvable tasks and 1 impossible task. It was set up carefully with a complex design. In the following, we explain the experimental design by focusing on the impossible task, which is the one we report on in this paper. Figure 1 gives an overview of the five steps.

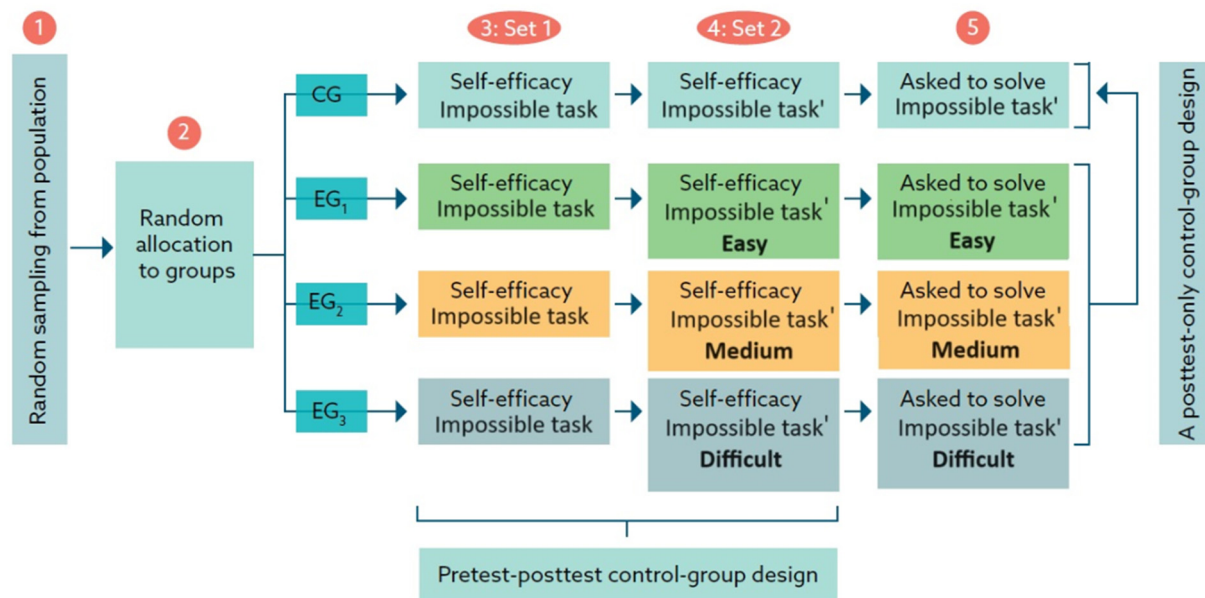


Figure 1. The Pretest-Posttest Control Group Design and Posttest-Only Control Group Design

In Step 1, we conducted a cluster sampling of participants across schools. In Step 2, the participants were randomly assigned (by a computer) to four groups: one control group (CG) and three experimental groups (EG_i, $i = 1, 2,$ and 3). Each group received different versions of the online survey, making this a true experiment. In Step 3 (Set 1),

the students were asked not to solve the task but to read the task and respond to the self-efficacy question—“How certain are you that you can solve this problem correctly?”—using a 100-point scale ranging from not certain at all (0) to absolutely certain (100), as recommended by the literature (Bandura, 2006; Zakariya et al., 2019). At this stage, the task had no level marking (see Figure 2), and the self-efficacy question was therefore used to measure what we referred to as the students’ “baseline self-efficacy.”

Since the purpose was to measure students’ self-efficacy and persistence, we further explain the survey’s *pretest-posttest control group* and *posttest-only control group* designs (Creswell & Creswell, 2018). In Step 4 (Set 2), the students were again presented with the task and asked the same self-efficacy question about how certain they were about being able to solve the task correctly. Here, too, the students were instructed not to solve the task. While the CG was given the same task all over again (with no level marking), for the students in the EGs, the task was now marked as easy, medium, or difficult. To avoid the feeling of having received the same task, as in Set 1, some words were replaced to make the tasks look different. Figure 2 shows how this played out for the task (e.g., “Casper” was replaced with “Emil” and “store” was replaced with “shop”). The reader is reminded that Sets 1 and 2 involved other tasks not included in the analysis for this paper, which made it less easy for the participating students to recognize the task at hand. The time interval between Step 3 and Step 4 was estimated to be approximately 20 minutes, as Step 3 was at the beginning of the survey and Step 4 was at the end of the survey.

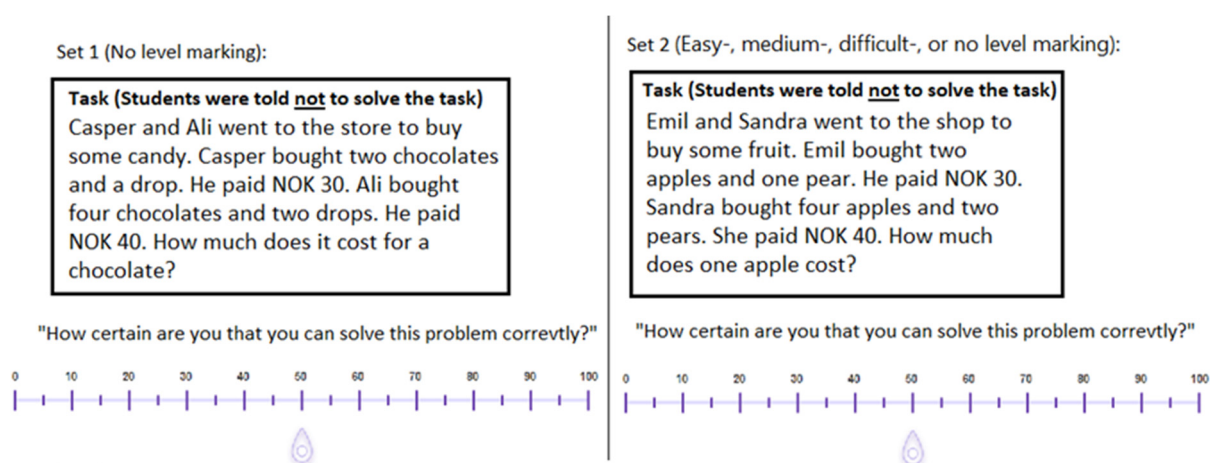


Figure 2. Measuring Self-Efficacy for the Specific Task in Sets 1 and 2

In Step 5, the students were told to solve the task and to write the answer in a text box. Persistence was measured by the time it took from a student was told to solve the task until he or she pressed the “send” button. The survey did not allow students to navigate between pages, making us more confident that the time measured was used only to solve the impossible task.

2.4 Statistical Analyses

We used the SPSS software package to test the hypotheses. First, the pretest-posttest control group design was used to measure the variable “difference in students’ self-efficacy,” which is the difference in students’ self-efficacy from Set 1 to Set 2 ($\text{self-efficacy}_{\text{Set2}} - \text{self-efficacy}_{\text{Set1}}$). This variable was used to analyze RQ₁ and RQ₂ by using a posttest-only control group design and was tested for normality using the method described by Kim (2013) and West et al. (1995). According to Kim (2013), when the sample size in each of the four groups was $50 \leq n < 300$, and z-values for skewness and kurtosis were less than ± 3.29 , we assumed normality. As our results met the assumption of normality, we used a series of independent samples t-tests to investigate how the level markings of the task affected differences in students’ self-efficacy by comparing the EGs’ results with the CG’s (RQ₁).

To differentiate between low and high “baseline self-efficacy,” we recoded the variable $\text{self-efficacy}_{\text{Set1}}$ into a two-level ordinal variable (low and high $\text{self-efficacy}_{\text{Set1}}$) using a median split, as suggested by Collins (1984). In this way, the participants in each of the four groups were divided into two roughly equal cell sizes, resulting in eight subgroups. Students’ “baseline self-efficacy” was recorded from $\text{self-efficacy}_{\text{Set1}}$ as follows: high baseline self-efficacy ($\text{self-efficacy}_{\text{Set1}} \geq 80$) and low baseline self-efficacy ($\text{self-efficacy}_{\text{Set1}} < 80$). To distinguish between the two subgroups, we marked each of the four groups (CG, EG₁, EG₂, and EG₃) with the letter “L” for low baseline self-efficacy and the letter “H” for high baseline self-efficacy (CG_L and CG_H, EG_{1L} and EG_{1H}, and so forth).

According to Kim (2013), when the sample size is $n < 50$ and the z-values for skewness and kurtosis are less than ± 1.96 , we can assume normality. The variable “difference in students’ self-efficacy” for students with low baseline self-efficacy in each of the four groups (CG_L, EG_{1L}, EG_{2L}, and EG_{3L}) met the normality assumption, and we therefore used a series of independent samples t-tests to investigate H_{2A}. For students with high baseline self-efficacy in three of the four groups (CG_H, EG_{2H}, and EG_{3H}), the variable met the normality assumption. However, for EG_{1H} the variable did not meet the normality assumption. Therefore, to investigate H_{2B}, we used a combination of independent samples t-tests and the Mann–Whitney U test.

In addition, we used a posttest-only control group design to analyze RQ₃ and RQ₄. As in previous research, we determined the maximum limit of time spent solving the task (Jacob et al., 1984) and recorded the variable “time” as either the amount of time spent solving the task or 20 min as a maximum. Since we could not assume normality, we used a Mann–Whitney U test to investigate the difference between students’ time spent solving the task for students in the EGs compared to students in the CG (H₃). In addition, because of non-parametric data, we used a series of Mann–Whitney U tests to examine how the interaction between the level marking of mathematics tasks and students’ baseline self-efficacy affected students’ persistence (H_{4A} and H_{4B}).

3. Results

The following section presents the findings of the study, which aimed at investigating how the level marking of a mathematics task affect students’ self-efficacy and persistence. Our RQs and hypotheses were formulated based on what the literature review revealed. First, we will present the results of testing H_{1A}, H_{1B}, H_{2A}, and H_{2B}, which all focused on self-efficacy, before we move on to the results of testing H₃, H_{4A}, and H_{4B} which focused on persistence. To measure the effect, the analysis focused on examining the difference between the CG and the EGs. The results contribute to a new understanding of how students’ self-efficacy and persistence can be influenced using level-marked tasks in tiered activities.

3.1 The Effects of Level Marking a Task on Students’ Self-Efficacy

To test H_{1A} and H_{1B}, we calculated the difference between students’ self-efficacy as measured in Set 1 (the task without a level marking) and in Set 2 (the task marked as easy, medium, or difficult or with no level marking). These differences are outlined in Table 1.

Table 1. Differences in Students’ Self-Efficacy in Sets 1 and 2 for All Four Groups

Group	<i>n</i>	Level marking of the task in Set 1	Students’ self-efficacy in Set 1 (M)*	Level marking of the task in Set 2	Students’ self-efficacy in Set 2 (M)	Difference in students’ self-efficacy (M)
CG	83	No marking	75.06	No marking	64.54	−10.52**
EG ₁	94	No marking	73.99	Easy	65.56	−8.43
EG ₂	68	No marking	72.79	Medium	64.84	−7.96
EG ₃	86	No marking	74.80	Difficult	57.03	−17.77

Note. *M = Mean. **The decrease in self-efficacy for this group shows the importance of having a CG. There can be changes in the level of self-efficacy during a test (see, e.g., Sexton & Tuckman, 1991).

As our results met the assumption of normality, Table 2 presents a series of independent samples t-tests (mean tests) used to examine H_{1A} and H_{1B}, where the difference in students’ self-efficacy between EG₁ and CG (Pair 1), EG₂ and CG (Pair 2), and EG₃ and CG (Pair 3) was tested. No significant differences were found, except in Pair 3, where there was a statistically significant decrease in students’ self-efficacy ($t = 1.795$, $df = 150$, $p = .037$, $d = 0.26$), with an effect size equal to 0.26 (a modest effect, according to Cohen, 2018).

Table 2. Results of Three Independent Samples T-Tests on the Differences in Students' Self-Efficacy

Pair**	Group	Level marking of the task in Set 2	<i>n</i>	Difference in students' self-efficacy (M)	Standard deviation (SD)	<i>t</i>	<i>df</i>	One-sided <i>p</i> -value
Pair 1	EG ₁	Easy	94	-8.43	20.39	.674	175	.251
	CG	No marking	83	-10.52	20.84			
Pair 2	EG ₂	Medium	68	-7.96	22.32	.728	149	.234
	CG	No marking	83	-10.52	20.84			
Pair 3	EG ₃	Difficult	86	-17.77	30.86	-1.795	150	.037*
	CG	No marking	83	-10.52	20.84			

Note. *The mean difference is significant at the .05 level ($p < .05$). **Levene's test for equality of variances (Cohen et al., 2018) was not significant for Pairs 1 and 2 ($p > .05$), and equal variances were assumed. In Pair 3 (EG3 and CG), equal variance was not assumed.

To test H_{2A} , we calculated the difference in self-efficacy as measured in Set 1 and Set 2 for students with low baseline self-efficacy (see Table 3).

Table 3. Difference in Self-Efficacy in Sets 1 and 2 for Students with Low Baseline Self-Efficacy for All Four Groups

Group	<i>n</i>	Level marking of the task in Set 1	Students' self-efficacy in Set 1 (M)	Level marking of the task in Set 2	Students' self-efficacy in Set 2 (M)	Difference in students' self-efficacy (M)
CG _L	49	No marking	51.68	No marking	45.42	-6.26
EG _{1L}	38	No marking	56.35	Easy	51.43	-4.92
EG _{2L}	33	No marking	51.73	Medium	47.39	-4.33
EG _{3L}	41	No marking	54.78	Difficult	52.46	-2.32

As our results met the assumption of normality, Table 4 shows a series of independent samples t-tests used to examine H_{2A} , in which the difference in students' self-efficacy between EG_{1L} and CG_L (Pair 1), EG_{2L} and CG_L (Pair 2), and EG_{3L} and CG_L (Pair 3) was tested. For students with low baseline self-efficacy, no significant change was found in their self-efficacy when an impossible task was marked by a difficulty level.

Table 4. Results of Three Independent Samples T-Tests on the Difference in Self-Efficacy for Students with Low Baseline Self-Efficacy

Pair*	Group	Level marking of the task in Set 2	<i>n</i>	Difference in students' self-efficacy (M)	Standard deviation (SD)	<i>t</i>	<i>df</i>	Two-sided <i>p</i> -value
Pair 1	EG _{1L}	Easy	49	-4.92	22.94	.254	85	.800
	CG _L	No marking	38	-6.26	26.40			
Pair 2	EG _{2L}	Medium	33	-4.33	26.08	.309	69	.758
	CG _L	No marking	38	-6.26	26.40			
Pair 3	EG _{3L}	Difficult	41	-2.32	26.13	.667	77	.507
	CG _L	No marking	38	-6.26	26.40			

Note. * Levene's test for equality of variances (Cohen et al., 2018) was not significant for Pairs 1-3 ($p > .05$), and equal variances were assumed.

In the same manner, the procedure used when testing H_{2A} was repeated for H_{2B} . The difference in self-efficacy as measured in Sets 1 and 2 for students with high baseline self-efficacy is presented in Table 5.

Table 5. Differences in Self-Efficacy in Relation to Different Level Markings for Students with High Baseline Self-Efficacy

Group	<i>n</i>	Level marking of the task in Set 1	Students' self-efficacy in Set 1 (M)	Level marking of the task in Set 2	Students' self-efficacy in Set 2 (M)	Difference in students' self-efficacy (M)
CG _H	45	No marking	94.80	No marking	80.69	-14.11
EG _{1H}	35	No marking	93.20	Easy	80.96	-12.24
EG _{2H}	45	No marking	92.66	Medium	81.29	-11.37
EG _{3H}	45	No marking	93.04	Difficult	61.20	-31.84

As our results met the assumption of normality, Table 6 shows the results of a series of independent samples *t*-tests used to examine H_{2A} , in which the difference in students' self-efficacy between EG_{2H} and CG_H (Pair 1) and EG_{3H} and CG_H (Pair 2) was compared. Interestingly, in Pair 2, there was a statistically significant decrease in students' self-efficacy ($t = 3.783$, $df = 88$, $p < .001$, $d = 0.80$), with an effect size equal to 0.80 (a large effect according to Cohen, 2018).

Table 6. Results of Two Independent Samples T-Tests on Differences in Self-Efficacy for Students with High Baseline Self-Efficacy

Pair*	Group	Level marking of the task in Set 2	<i>n</i>	Difference in students' self-efficacy (M)	Standard deviation (SD)	<i>t</i>	<i>df</i>	Two-sided <i>p</i> value
Pair 1	EG _{2H}	Medium	35	-11.37	17.81	.772	78	.443
	CG _H	No marking	45	-14.11	13.95			
Pair 2	EG _{3H}	Difficult	45	-31.84	18.18	-3.783	88	< .001**
	CG _H	No marking	45	-14.11	13.95			

Note. *Levene's test for equality of variances (Cohen et al., 2018) was not significant for Pair 1 ($p > .05$), and equal variance was assumed. For Pair 2 (EG_{3H} and CG_H), equal variance was not assumed. ** The mean difference was significant at the .05 level ($p < .05$)

The difference in students' self-efficacy in EG_{1H} did not meet the normality assumption, and we therefore utilized a Mann-Whitney U test (a median test) to compare students' median of difference in self-efficacy between EG_{1H} and CG_H. As shown in Table 7, the median (Med) in EG_{1H} (Med = 5.0) was not statistically significantly different compared to CG_H (Med = 10.0, $p = .379$).

Table 7. Results of One Mann-Whitney U Test on the Difference in Self-Efficacy for Students with High Baseline Self-Efficacy

Group	Level marking	<i>n</i>	Med	<i>U</i>	<i>Z</i>	Two-sided <i>p</i> -value
EG _{1H}	Easy	45	5.0	904.500	-.881	.379
CG _H	No marking	45	10.0			

Note. Med: Median

3.2 The Effects of Level Marking on Students' Persistence

To test H_3 , we investigated how "easy," "medium," and "difficult" markings affected students' persistence by comparing the students' time spent on tasks without level markings (CG) and with level markings (EGs). The median time (in seconds) for all students ($n = 331$) is illustrated in Table 8.

Table 8. Results of Three Mann–Whitney U Tests on Students’ Time Spent on Tasks

Group	Level marking	<i>n</i>	Med	<i>U</i>	<i>Z</i>	Two-sided <i>p</i> -value
EG ₁	Easy	94	120.5	3709	-.563	.573
CG	No marking	83	100.0			
EG ₂	Medium	68	99.0	2597	-.842	.400
CG	No marking	83	100.0			
EG ₃	Difficult	86	55.5	2825	-2.340	.019*
CG	No marking	83	100.0			

Note. Med: Median. *The difference was significant at the .05 level.

As shown in Table 8, the Mann–Whitney U tests revealed that the time spent solving a task was not significantly different for tasks with “easy” and “medium” markings compared to the CG. Interestingly, one Mann–Whitney U test revealed significant differences in students’ time spent on the task marked “difficult” (Med = 55.5, *n* = 86) and one with no level marking (Med = 100.0, *n* = 83, *U* = 2825, *Z* = -2.340, *p* = .019, *r* = 0.18), with an effect size equal to 0.18 (a small effect according to Cohen, 2018).

To test H_{4A}, we investigated how “easy,” “medium,” and “difficult” markings affected students’ persistence by comparing the time spent on tasks without a level marking (CG_L) and with level markings (EG_{SL}) for students with low baseline self-efficacy. For students with low baseline self-efficacy (*n* = 161), the median time (in seconds) they spent trying to solve a task with different level markings is shown in Table 9.

Table 9. Results of Three Mann–Whitney U Tests on Time Spent on Tasks among Students with Low Baseline Self-Efficacy

Group	Level marking	<i>n</i>	Med	<i>U</i>	<i>Z</i>	Two-sided <i>p</i> -value
EG _{1L}	Easy	49	77.0	846.00	-.727	.467
CG _L	No level marking	38	68.0			
EG _{2L}	Medium	33	93.0	621.00	-.069	.945
CG _L	No level marking	38	68.0			
EG _{3L}	Difficult	41	52.0	675.50	-1.016	.310
CG _L	No level marking	38	68.0			

Note. Med: Median

As shown in Table 9, the Mann–Whitney U test for students with low baseline self-efficacy revealed no significant differences in students’ time spent on a task with any of the level markings (EG_{SL}) compared to no level marking (CG_L).

To test H_{4B}, we investigated how the different level markings affected persistence by comparing time spent on the task without a level marking (CG_H) and with level markings (EG_{SH}). For students with high baseline self-efficacy, the median time (in seconds) spent solving tasks with different level markings is shown in Table 10.

Table 10. Results of Three Mann–Whitney U Tests on Time Spent on Tasks among Students with High Baseline Self-Efficacy

Group	Level marking	<i>N</i>	Med	<i>U</i>	<i>Z</i>	Two-sided <i>p</i> -value
EG _{1H}	Easy	45	134.00	978.500	-.274	.784
CG _H	No level marking	45	144.00			
EG _{2H}	Medium	35	105.00	669.500	-1.145	.252
CG _H	No level marking	45	144.00			
EG _{3H}	Difficult	45	72.00	742.00	-2.184	.029*
CG _H	No level marking	45	144.00			

Note. Med: Median. *The difference was significant at the .05 level.

The Mann–Whitney U tests for students with high baseline self-efficacy revealed that the time spent solving a task was not significantly different for easy- and medium-level markings compared to the control group. The Mann–Whitney U test for students with high baseline self-efficacy revealed a significant difference in students' time spent on a task with a “difficult” marking (Med = 72.00, $n = 45$) compared to no level marking (Med = 144.0, $n = 45$, $U = 742.00$, $Z = -2.184$, $p = .029$, $r = 0.23$), with an effect size equal to 0.23 (a small effect according to Cohen, 2018).

4. Discussion

Level-marked tasks is used in tiered activities (e.g., Brändström, 2005; Grave & Pepin, 2015) to facility mastery experiences. However, little research has been conducted to investigate how the difficulty labels contributes in the tiered activities to enhancing students' self-efficacy and persistence. The aim of this study was to investigate how the level marking of mathematics tasks affects students' persistence and self-efficacy, revealing discouraging results that, in isolation, do not speak in favor of using level-marked tasks in tiered activities. Our findings revealed that none of the labels (easy, medium and difficult) increased students' self-efficacy and persistence. In fact, the results indicated that difficult level marking of the tasks reduced both students' mathematics self-efficacy and their persistence. In this section, we discuss our results, organized by our four RQs.

4.1 Level Marking of Mathematics Tasks and Student's Self-Efficacy

In RQ₁, we asked how the level marking of a mathematics task that is impossible to solve affects students' self-efficacy. We found that when the task that we reported on here was marked as difficult, it had a significantly negative effect on students' self-efficacy. This corroborates what has been reported elsewhere (Herset et al., 2023)—marking an easy task as difficult has a negative effect on students' self-efficacy, revealing that students tend not to judge their ability to solve a task by its content but rather by its level marking.

The results presented by Herset et al. (2023) and the ones we report on here are in good agreement with Bandura (1997) and Street et al. (2017, 2022), who claimed that a person's level of self-efficacy is influenced by their opinion of the difficulty of a task, as well as Liu et al. (2020), who suggested that students' task-specific self-efficacy can be affected by perceived task difficulty and not by assessments of their own capabilities. However, the current study does not support a positive effect of “easy” or “medium” markings of a task on students' self-efficacy, a result that diverges from previous research reporting that students' self-efficacy was higher when they were told that a task was easy compared to when they were told that a task was difficult (Jacobs et al., 1984). That being said, it is important to point out that the study carried out by Jacobs et al. (1984) did not have a control group, and their findings were also limited by the absence of a pretest. One interpretation of the results, which have implications for theory and practice in mathematics education, is that marking a task as “easy” is not necessary enough to increase students' self-efficacy. This is justified by the fact that students' previous experiences can influence their self-efficacy, and if they experience that tasks marked as “easy” are not always easy to solve, the students may become skeptical of easy marking and thus maintain their original self-efficacy.

In RQ₂, considering students' baseline self-efficacy, we asked how the level marking of a mathematics task that is impossible to solve affect students' self-efficacy, which made us set forth two hypotheses. First, regarding students with low baseline self-efficacy, no significant change was found in their self-efficacy after being presented with the level marking of an impossible task, not even when the task was marked as difficult. From this, we infer those students with low self-efficacy in Set 1 considered the task difficult, and when the task was marked as difficult (in Set 2), their perception of the task as difficult did not change. In contrast, for students with high baseline self-efficacy, we found that marking the task as difficult resulted in significant reductions in self-efficacy. This finding is a new contribution to the research field, as the effect seems to be particularly prominent among students with high baseline self-efficacy. Moreover, we found that marking a task as “easy” or “medium” did not affect these students' self-efficacy. It is reasonable to conclude that students with high self-efficacy in Set 1 may have assessed the level of the task (without a level marking) as easy or moderately difficult, and when the tasks were marked as “easy” or “medium” (in Set 2), this did not affect their self-efficacy. Taken together, our results resonate well with the results of Collins (1984), who found that students with high self-efficacy rated tasks as significantly easier than students with low self-efficacy.

4.2 Level Marking of Mathematics Tasks and Students' Persistence

In RQ₃, we asked how the level marking of a mathematics task that is impossible to solve affects students' persistence, and we found that the “difficult” marking had a negative effect. That is, when students were faced with a “difficult” marking, the time they spent trying to solve the task changed, a result that is consistent with those of Herset and El Ghami (2022), who found a negative effect of marking a hard (but solvable) task as difficult on the time students spent trying to solve it. However, as in the case of self-efficacy (RQ₁), the current study does not

support the effect of “easy” or “medium” markings on students’ persistence. This means that there may be a connection between students’ self-efficacy and the time spent on an impossible task. This is justified by previous research that found a positive relationship between self-efficacy and persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Schunk, 1991; Sexton & Tuckman, 1991; Skaalvik et al., 2015), which underlines how important self-efficacy might be on persistence. The findings we report here add more to the findings of Street et al. (2017), who suggested that students’ perceived task difficulty affects self-efficacy.

In RQ4, considering students’ baseline self-efficacy, we asked how the level marking of a mathematics task that is impossible to solve affect students’ persistence, which made us set forth two hypotheses. First, regarding students with low baseline self-efficacy, no significant change was found in their persistence following encountering the level marking of an impossible task, not even when the task was marked as difficult. Therefore, we argue that the level markings for students with low baseline self-efficacy does not affect persistence. This may be because their opinions of the level of difficulty of the task were the same in Sets 1 and 2, regardless of the level marking. Another explanation could be that students with low baseline self-efficacy lacked mastery experience during the survey. When participants were asked to solve the last impossible task, they had already been given other solvable tasks to complete; thus, their mastery experience may have influenced their persistence. This is in good agreement with Sexton and Tuckman (1991), who found that at the beginning of solving a task, generalized self-efficacy and past performance were the best predictors of persistence, but after solving a few tasks, it was students’ specific self-efficacy and how much time they had spent on the test previously that best predicted persistence. Furthermore, those with high baseline self-efficacy spent significantly less time trying to solve the task when the task was marked as difficult. The result fits well with Jacobs et al. (1984), who found that students who were told that the task was difficult reduced their self-efficacy and persistence compared to students’ who were told the opposite. However, our findings suggest something new; this effect is particularly prominent among students with high baseline self-efficacy.

4.3 Limitations

We acknowledge the limitations of this study, which are mainly related to the lack of pretesting when investigating students’ persistence. In addition, persistence was operationalized as the students’ time spent solving an impossible task. We therefore do not know if the amount of time from receiving the task to when they pushed the “send” button was used to solve the task or if they were doing something else (e.g., some students may start to daydream). We therefore acknowledge potential alternative interpretations of the study, like for instance that the students may have felt the need to hurry because the school day was almost over. To overcome such limitations, we suggest observation of students’ during the survey, to examine how students are spending their time when working with mathematics tasks with varying level markings. Moreover, interviews can provide insight into why students spend less time solving a task marked as difficult. This issue regarding the lack of knowledge about students’ actions while solving the task was also raised by DiNapoli (2023) and is one of the reasons why we included a control group. Since the size of the sample is relatively large ($n = 331$) and the allocation to the CG and EGs was random, it is plausible that a random measurement error was mitigated.

Limitations in the data collection process include the sampling strategy, sample size, and the skewing in the distribution of participants. Because of COVID-19, we had to choose several schools in one district to receive enough data. Based on the table of sample sizes, a sample size of 383 students was recommended (Cohen et al., 2018). This means that our full sample was above the limit ($n = 436$), but unfortunately, due to the missing values, outliers (as described above), and the way in which the students were divided into subgroups, the sample size may be a limitation.

4.3 Implications and Concluding Remarks

In the context of learning mathematics, both self-efficacy and persistence are important to gain mastery experience. We know from previous research that self-efficacy influences persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Schunk, 1991; Skaalvik et al., 2015), and persistence is important in getting opportunities to learn (Peterman & Seligman, 2004). Knowing this, and further acknowledging that mastery experiences affect both self-efficacy (e.g., Butz & Usher, 2015; Joët et al., 2011; Stevens et al., 2006; Usher & Pajares, 2009; Zientek et al., 2019) and persistence (Bong & Skaalvik, 2003; Collins, 1984; Jacobs et al., 1984; Lyman et al., 1984; Multon et al., 1991; Schunk, 1991; Skaalvik et al., 2015), we assert that differentiated instruction that results in mastery are important—such as a tiered activity according to students readiness levels.

However, the results from this study have, to some extent, enhanced our understanding of the use of level-marked tasks in a tiered activity. Our findings suggest that the use of level-marked tasks does not contribute to higher self-efficacy and persistence in mathematics, which is contrary to the intended goal of tiered activities aimed to help students achieve an optimal learning effect and increase self-efficacy. If “difficult” markers affect students’ persistence, meaning that students spend less time solving a task, they may affect students’ learning outcomes. This is supported by Street et al. (2022), who recommended supporting students’ engagement when solving challenging tasks, as this will strengthen self-efficacy and performance. Our findings suggest that labeling tasks as difficult is not a good way to support students’ engagement, as the results suggest that students with high self-efficacy experience lowered self-efficacy and spend less time trying to solve tasks when they are informed that the task is difficult. Because previous research found tiered activity be significant in mathematics learning (Bal, 2016; Luster, 2008; Scott, 2012; Suarez, 2007), we still believe that the activity is important in mathematics. However, in this study, we find that the use of level-marked tasks in tiered activities seemingly do not contribute to increase students’ self-efficacy and persistence. In addition, as we were able to compare the results from the EGs with the CG, the study unveiled novel insight: informing students that the task is easy does not increase students’ self-efficacy and persistence, which is contrary to the findings of Jacobs et al. (1984). Hence, we suggest that the research we report on here has some methodological strengths that shed some new lights to the 40-year-old results of Jacobs et al. (1984).

Based on our results, we suggest that caution should be exercised in using level-marked tasks as a tool in differentiated instruction. Marking tasks as easy, medium and difficult alone does seem to be insufficient for a tiered activity and cannot be considered a guaranteed solution for enhancing students’ self-efficacy and persistence. We therefore recommend mathematics teachers to support students when solving tasks marked as difficult, by for instance expressing faith in students’ capabilities. In addition, we suggest that teachers inform students that the level of difficulty is a subjective evaluation (Krauthausen, 2018) and that the level marking of a task may not always be correct (Brändström, 2005; Singh, 2017). However, more research is needed to confirm the findings we report on in this paper, using both other tasks and other student groups. In addition, we recommended that further research examine how our results are related to other variables (e.g. students’ long-term goals and motivation in mathematics).

References

- Auliya, K., & Widjajanti, D. B. (2023). Singaporean and Japanese maths textbooks: Character, structure, and content. *Mosharafa: Jurnal Pendidikan Matematika*, 12(1), 155-168. <https://doi.org/10.31980/mosharafa.v12i1.2141>
- Bal, A. P. (2016). The effect of the differentiated teaching approach in the algebraic learning field on students’ academic achievements. *Eurasian Journal of Educational Research*, 16(63), 185-204. <https://doi.org/10.14689/ejer.2016.63.11>
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117-148. https://doi.org/10.1207/s15326985ep2802_3
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Bandura, A. (2006). Guide to the construction of self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (vol. 5, pp. 307-337). New York: Information Age.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1-40. <https://doi.org/10.1023/A:1021302408382>
- Borge, A. I. H. (2003). Psykologi og forskningsetikk: Kan deltagelse i forskningsprosjekt gi psykiske skader? [Psychology and research ethics: Can participation in a research project cause psychological damage?] I K. Ruyter (Ed.), *Forskningsetikk–Beskyttelse av enkeltpersoner og samfunn* (pp. 93-107). Oslo: Gyldendal Akademisk.
- Brändström, A. (2005). *Differentiated tasks in mathematics textbooks. An analysis of the levels of difficulty*. [Unpublished licentiate thesis]. Luleå University of Technology.
- Brown, I., & Inouye, D. K. (1978). Learned helplessness through modeling: The role of perceived similarity in competence. *Journal of Personality and Social Psychology*, 36(8), 900-908. <https://doi.org/10.1037/0022-3514.36.8.900>
- Butz, A. R., & Usher, E. L. (2015). Salient sources of early adolescents’ self-efficacy in two domains. *Contemporary Educational Psychology*, 42, 49-61. <https://doi.org/10.1016/j.cedpsych.2015.04.001>

- Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education*, 75(3), 221-244. <https://doi.org/10.3200/JEXE.75.3.221-244>
- Cohen, L., Manion, L., & Morrison, A. K. (2018). *Research methods in education*. London: Routledge.
- Collins, J. L. (1984). *Self-efficacy and ability in achievement behavior* [Doctoral dissertation, Stanford University].
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Quantitative, qualitative and mixed methods*. California: Sage.
- DiNapoli, J. (2023). Distinguishing between grit, persistence, and perseverance for learning mathematics with understanding. *Education Sciences*, 13(4), 402. <https://doi.org/https://doi.org/10.3390/educsci13040402>
- Grave, I., & Pepin, B. (2015). Teachers' use of resources in and for mathematics teaching. *Nordic Studies in Mathematics Education*, 20(3-4), 199-222.
- Eriksen, E., Solomon, Y., Bjerke, A. H., Gray, J., & Kleve, B. (2022). Making decisions about attainment grouping in mathematics: Teacher agency and autonomy in Norway. *Research Papers in Education*, 1-21. <https://doi.org/10.1080/02671522.2022.2135014>
- Hackett, G., and Betz, N. E. (1989). An Exploration of the Mathematics Self-Efficacy/Mathematics Performance Correspondence. *Journal for Research in Mathematics Education*, 20(3), 261-273. <https://doi.org/10.2307/749515>
- Herset, M., & El Ghami, M. (2022). The effect of level-marking mathematical tasks on students' time spent on such tasks and correct solutions: An experimental study. *Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*, Feb 2022, Bozen-Bol, Italy.
- Herset, M., El Ghami, M., & Bjerke, A. H. (2023). The effect of level-marked mathematics tasks on students' self-efficacy: An experimental study. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1116386>
- Jacobs, B., Prentice-Dunn, S., & Rogers, R. W. (1984). Understanding persistence: An interface of control theory and self-efficacy theory. *Basic and Applied Social Psychology*, 5(4), 333-347. https://doi.org/10.1207/s15324834basp0504_6
- Joët, G., Usher, E. L., & Bressoux, P. (2011). Sources of self-efficacy: An investigation of elementary school students in France. *Journal of Educational Psychology*, 103(3), 649. <https://doi.org/10.1037/a0024048>
- Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52-54. <https://doi.org/10.5395/rde.2013.38.1.52>
- Krauthausen, G. (2018). Natural differentiation – An approach to cope with heterogeneity. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Hrsg.), *Invited lectures from the 13th International Congress on Mathematical Education*. ICME-13 Monographs. Springer Open. https://doi.org/10.1007/978-3-319-72170-5_19.
- Liu, Q., Liu, J., Cai, J., & Zhang, Z. (2020). The relationship between domain- and task-specific self-efficacy and mathematical problem posing: A large-scale study of eighth-grade students in China. *Educational Studies in Mathematics*, 105(3), 407-431. <https://doi.org/10.1007/s10649-020-09977-w>
- Luster, R. (2008). A quantitative study investigating the effects of whole-class and differentiated instruction on student achievement. [Doctoral dissertation, Walden University].
- Lyman, R. D., Prentice - Dunn, S., Wilson, D. R., & Bonfilio, S. A. (1984). The effect of success or failure on self-efficacy and task persistence of conduct - disordered children. *Psychology in the Schools*, 21(4), 516-519. [https://doi.org/10.1002/1520-6807\(198410\)21:4<516::AIDPITS2310210419>3.0.CO;2-O](https://doi.org/10.1002/1520-6807(198410)21:4<516::AIDPITS2310210419>3.0.CO;2-O)
- Miele, D. B., Browman, A. S., Shen, C., Vasilyeva, M. & Tyumeneva, Y. A. (2022). Domain-general and math-specific self-perceptions of perseverance as predictors of behavioral math persistence. *The Journal of Experimental Education*, 90(3), 593-614. <https://doi.org/10.1080/00220973.2020.1799312>
- Montague, M., & Applegate, B. (2000). Middle school students' perceptions, persistence, and performance in mathematical problem solving. *Learning Disability Quarterly*, 23(3), 215-227. <https://doi.org/10.2307/1511165>
- Moyer, J. C., Robison, V., & Cai, J. (2018). Attitudes of high-school students taught using traditional and reform mathematics curricula in middle school: A retrospective analysis. *Educational Studies in Mathematics*, 98, 115-134. <https://doi.org/10.1007/s10649-018-9809-4>

- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 38(1), 30-38. <https://doi.org/10.1037/0022-0167.38.1.30>
- Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. Snyder & S. Lopez (Ed.), *Handbook of positive psychology* (pp. 89-105). New York: Oxford University Press.
- Óturai, G., Riener, C., & Martiny, S. E. (2023). Attitudes towards mathematics, achievement, and drop-out intentions among STEM and non-STEM students in Norway. *International Journal of Educational Research Open*, 4. <https://doi.org/10.1016/j.ijedro.2023.100230>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-478. <https://doi.org/10.2307/1170653>
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, 42(2), 190-198. <https://doi.org/10.1037/0022-0167.42.2.190>
- Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.
- Pierce, R. L., & Adams, C. M. (2005). Using tiered lessons in mathematics. *Mathematics Teaching in the Middle School*, 11(3), 144-149. <https://doi.org/10.5951/MTMS.11.3.0144>
- Sasidharan, S., & Kareem, J. (2023). Mathematics self-efficacy, utility value and well-being among school students in India mediating role of student engagement. *Investigations in Mathematics Learning*, 1-13. <https://doi.org/10.1080/19477503.2023.2224652>
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26(3-4), 207-231. <https://doi.org/10.1080/00461520.1991.9653133>
- Scott, B. E. (2012). *The effectiveness of differentiated instruction in the elementary mathematics classroom* [Doctoral dissertation, Ball State University, Muncie, Indiana].
- Sexton, T. L., & Tuckman, B. W. (1991). Self-beliefs and behavior: The role of self-efficacy and outcome expectation over time. *Personality and Individual Differences*, 12(7), 725-736. [https://doi.org/10.1016/0191-8869\(91\)90228-4](https://doi.org/10.1016/0191-8869(91)90228-4)
- Shen, C., Miele, D. B., & Vasilyeva, M. (2016). The relation between college students' academic mindsets and their persistence during math problem solving. *Psychology in Russia: State of the Art*, 9(3), 38-56. <https://doi.org/10.11621/pir.2016.0303>
- Skaalvik, E. M., Federici, R. A., & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research*, 72, 129-136. <https://doi.org/10.1016/j.ijer.2015.06.008>
- Spielberg, Y., & Azaria, A. (2021). Revelation of task difficulty in AI-aided education. In *2021 IEEE 33rd International conference on tools with artificial intelligence (ICTAI)* (pp. 1403-1408). IEEE. <https://doi.org/10.1109/ICTAI52525.2021.00223>
- Stein, M., & Burchartz, B. (2006). The invisible wall project: Reasoning and problem solving processes of primary and lower secondary students. *Mathematical Thinking and Learning*, 8(1), 65-90. https://doi.org/10.1207/s15327833mtl0801_4
- Stevens, T., Olivárez Jr., A., & Hamman, D. (2006). The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students. *Hispanic Journal of Behavioral Sciences*, 28(2), 161-186. <https://doi.org/10.1177/0739986305286103>
- Street, K. E. S., Malmberg, L.-E., & Stylianides, G. J. (2017). Level, strength, and facet-specific self-efficacy in mathematics test performance. *ZDM: Mathematics Education*, 49(3), 379-395. <https://doi.org/10.1007/s11858-017-0833-0>
- Street, K. E., Stylianides, G. J., & Malmberg, L. E. (2022). Differential relationships between mathematics self-efficacy and national test performance according to perceived task difficulty. *Assessment in Education: Principles, Policy & Practice*, 29(3), 288-309. <https://doi.org/10.1080/0969594X.2022.2095980>
- Suarez D. (2007). Differentiation by challenge: Using a tiered program of instruction in mathematics. In W. Powel & O. K. Powel (Eds.), *Making the difference: Differentiation in international schools* (pp. 199-227). Kuala

Lumpur: EAF Press.

- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Alexandria: ASCD.
- Tomlinson, C. A. (2005). Differentiating instruction: Why bother? *Middle Ground*, 9(1), 12-14.
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89-101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). London: Sage.
- Zakariya, Y. F. (2019). Study approaches in higher education mathematics: investigating the statistical behaviour of an instrument translated into Norwegian. *Education sciences*, 9(3), 191. <https://doi.org/10.3390/educsci9030191>
- Zakariya, Y. F. (2021). Self-efficacy between previous and current mathematics performance of undergraduate students: An instrumental variable approach to exposing a causal relationship. *Frontiers in Psychology*, 11, 1-11. <https://doi.org/10.3389/fpsyg.2020.556607>
- Zientek, L. R., Fong, C. J., & Phelps, J. M. (2019). Sources of self-efficacy of community college students enrolled in developmental mathematics. *Journal of Further and Higher Education*, 43(2), 183-200. <https://doi.org/10.1080/0309877X.2017.1357071>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).