



Undergraduate students' attitudes towards mathematical proving in an introduction to proof course

Jokke Häsä^{1,2} · Lín Westlin³ · Johanna Rämö⁴

Accepted: 2 May 2023 / Published online: 15 July 2023
© The Author(s) 2023

Abstract

In this study, we investigated undergraduate mathematics students' ($N=267$) attitudes towards proving. The students were taking an introduction-to-proof type course that was situated at the beginning of the mathematics curriculum and lasted for one term. Four attitude variables were measured at the beginning and at the end of the course with a new self-report instrument: self-efficacy, anxiety, appreciation and motivation. The instrument was based on two existing instruments on mathematics attitudes and proof-related self-efficacy. We studied how these four attitude variables were related to the students' prior skills and their gender at the beginning of the course, how the attitude variables changed during the course, and how they affected the students' performance in the final course project. Our results indicate that students' prior performance is linked to their proof-related self-efficacy, anxiety and motivation at the beginning of the course. Female students exhibited lower efficacy and motivation levels than male students. During the focus course, students' self-efficacy increased and anxiety decreased. The gender gaps in self-efficacy and motivation persisted throughout the course. In addition, high motivation in the beginning of the course predicted good performance in the final project. Based on our results, we conclude that an introductory course on proving can enhance students' attitudes, and we suggest that these attitudes are taken into account in teaching as they can affect students' performance. Finally, we urge researchers and professionals to earnestly consider ways to mitigate gender differences in mathematics.

Keywords Mathematical proof · Attitudes · Self-efficacy · Gender differences · Introduction to proof

✉ Jokke Häsä
jokke.hasa@helsinki.fi

¹ Department of Education, Centre for University Teaching and Learning, University of Helsinki, Helsinki, Finland

² Present Address: Finnish Institute for Health and Welfare (THL), Helsinki, Finland

³ Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

⁴ Department of Physics and Mathematics, University of Eastern Finland, Joensuu, Finland

1 Introduction

Proving has a central role in mathematical practices. Proofs are used to verify mathematical results, deduce new results from existing facts and communicate one's reasoning to other mathematicians. In mathematics education, they can also be used to convey to students important mathematical methods and strategies (Hanna & Barbeau, 2010). For these reasons, university mathematics education generally emphasises proving. However, in many countries, proving is rarely studied in school, and even then only in specific contexts, such as geometry (see e.g., Furinghetti et al., 2011; Mingus & Grassl, 1999). Not surprisingly, students find proof one of the most difficult areas of mathematics (Moore, 1994), and the transition to proving when entering university from secondary education is particularly challenging for them (Selden, 2012). Mathematical proving has been studied widely and from many perspectives such as proof construction, reading and validation (e.g., Alcock & Weber, 2010; Hodds et al., 2014; Selden & Selden, 2009). Also, students' beliefs about the role and nature of proof have been investigated (e.g., Almeida, 2000; Conner et al., 2011), as well as pre-service teachers' views on teaching of proof (e.g., Lesseig & Hine, 2022; Schwarz et al., 2008). Our standpoint is different from the aforementioned studies. In this study, our focus is on certain affective dimensions related to proving, each of which we consider an attitude towards proving. We chose this focus as the study of affective attributes in mathematical proving seems to be an underrepresented point of view in the research literature, although they play an important role in other mathematics education research.

There is no single definition of attitudes in the literature. In this work, we consider as *attitude* any persistent affective disposition or belief that may evoke emotions and guide behaviour (cf. Leder, 1987; McLeod, 1992). Studying such attitudes, including enjoyment, appreciation and self-confidence, has a long history in mathematics education research (Di Martino & Zan, 2015; Hannula, 2002; Leder, 1987; McLeod, 1992). Attitudes have been considered important predictors of study behaviour, and hence also of achievement in mathematics (e.g., Singh et al., 2002). For example, students' beliefs in their own capabilities have a positive effect on achievement in mathematics (Pajares & Kranzler, 1995), as does considering mathematics useful and valuable (Greene et al., 1999).

It is also known that attitudes towards mathematics differ between genders. Female students have been observed to have lower self-confidence in mathematics, to suffer from more mathematics anxiety than male students and to be more poorly motivated to study mathematics than male students (Frost et al., 1994; Hembree, 1990; Middleton & Spanias, 1999). These differences are known to hinder female students' performance and are suspected to contribute to the underrepresentation of women in science careers (Betz & Hackett, 1983; Frost et al., 1994).

In this study, we focus on proving and on four attitudinal variables that we have assumed to play a significant role in learning proving. These are self-efficacy, anxiety, appreciation and motivation. Being motivated, believing in one's own abilities and finding value in the topic of study can assist in overcoming the mental challenges of the transition to proof. However, there might be gender differences in these attitudinal variables. We will justify our assumptions by presenting examples from existing literature on attitudes towards mathematics. We will then explore the research gap by studying relationships between attitudes, gender and performance during the transition to university mathematics and proving.

2 Aim of the study

We investigated students' attitudes towards proving in a 14-week undergraduate mathematics course. The purpose was to explore how students' attitudes towards proving change during such a course, and how they might influence achievement. We also investigated what role prior performance and gender play in this interaction. This initial study focuses on one teaching context, paving the way for more extensive studies.

To measure the attitudes, we combined two existing questionnaires: one measuring attitudes towards mathematics (ATMI; Tapia, 1996) and one measuring proof self-efficacy (PSE; Iannone & Inglis, 2010). The first was chosen because it was rather short and contained items related to confidence, anxiety, value, enjoyment, and motivation, which we considered important in the context of proving. The second was chosen because it focused especially on proving and might therefore strengthen the validity of the full instrument. Based on these questionnaires, we developed a new instrument that measured four attitude variables: self-efficacy, anxiety, appreciation and motivation.

First, we were interested in the attitudes that students have when they come to university. It is known that previously acquired skills are correlated with self-efficacy beliefs and anxiety towards a subject (Sitzmann & Yeo, 2013; Ma & Kishor, 1997; Ho et al., 2000; Miller & Bichsel, 2004), as well as motivation (Middleton & Spanias, 1999; Murayama et al., 2013; Singh et al., 2002). It is also known that there are gender differences in attitudes towards mathematics (e.g., Frost et al., 1994; Lahdenperä, 2018; Leder, 1995; Middleton & Spanias, 1999; Wigfield & Meece, 1988), so it is likely that these differences would be found also in attitudes towards proving. This leads to our first research question:

RQ1. How are students' attitudes towards proving related to their gender and high school mathematics performance at the beginning of university studies?

Next, the focus course has been designed to smooth the transition to university mathematics, and a large part of the course is dedicated to teaching mathematical proving. It would be expected that this kind of course might have a positive effect on students' attitudes towards proving. It is also possible that the learning environment may either increase or decrease a gender gap in attitude variables. To investigate this, we compared pre- and post-course surveys to answer the following research question:

RQ2. How did students' attitudes towards proving change during the course, and what effect did gender have on the change?

Finally, we wanted to explore to what extent attitudes can have a causal effect on proving performance. To this end, we examined the students' performance in the final project of the focus course, comparing it with the attitude variables measured at the beginning of the course and controlling for high school mathematics performance. As the final project included tasks related to proving, we aimed to answer the following question:

RQ3. How do differences in attitudes towards proving affect performance in proof-related tasks?

In the final question, we also distinguished between male and female students.

3 Theoretical background

3.1 Attitudes and their measurement in mathematics education

By attitudes in the context of learning, we refer to learned predispositions that evoke emotional responses towards the subject one is learning, involve beliefs and values about the subject, may guide the learner's behaviour and are fairly persistent (cf. Leder, 1987; McLeod, 1992). For example, a learner can dislike geometry, find algebra boring, consider mathematics important or themselves good at it. Attitude is not the same as emotion, since emotions can be temporary (e.g., feeling sad after a failed exam). Attitudes are also not identical to beliefs, in that beliefs can be completely unemotional (e.g., a belief that mathematics is needed in engineering). Typically, attitudes are classified as positive or negative, and sometimes learners are described simply as "having a positive or negative attitude" towards the subject. We, however, consider attitude as a multidimensional construct.

Historically, attitudes are situated in the affective domain. The affective domain, or affect, refers to those psychological processes that reside outside cognition. Regarding affect in mathematics education, McLeod (1992) considered a tripartite model consisting of beliefs, attitudes and emotions that would play a major role in organising the affective domain. Later, DeBellis and Goldin (2006) added values to this model. Although the model is generally considered valuable and important, it has been difficult to formulate clear, separate definitions for the individual concepts (Di Martino & Zan, 2011). Attitudes, in particular, have been difficult to give a comprehensive definition (Leder, 1987). One theoretical characterisation of attitudes was offered by Hannula (2002), who described attitude as "a category of behaviour that is produced by different evaluative processes." These evaluative processes would occur in various learning situations, and would be directed by emotions, expectations and values. On a different note, Di Martino and Zan (2010) used a grounded theory approach to students' narratives in an attempt to distil a pragmatic definition of attitude. They produced a three-dimensional model consisting of emotional disposition such as like or dislike, vision of mathematics including value and usefulness, and perceived competence regarding beliefs about self.

Apart from theoretical investigations, there have been many attempts at measuring attitudes towards mathematics, and many kinds of self-report instruments have been developed for this purpose (e.g., Di Martino & Zan, 2015; Leder, 1985). The interest towards measurement may have stemmed from the idea that attitudes predict achievement (Aiken, 1970, as cited in Di Martino & Zan, 2015), although the meta-analysis by Ma and Kishor (1997) found the overall effect size between the two to be too small to have meaningful implications for educational practice. It can be noted that the attitudes measured with self-report instruments often include beliefs, values, emotions and other constructs that may not reflect a particular theoretical definition of attitude. In fact, Leder (1985) notes that research conducted with these instruments typically uses a definition of attitude derived from what is measured. He states that this may stem from the difficulty to match the conceptual and operational definitions. The measurements may still provide valuable information "provided the measures used are both valid and reliable," reflecting an "appropriate conceptualisation of attitude" (Leder, 1985, p. 19).

An early and influential instrument was produced by Fennema and Sherman (1976), measuring attitudes towards mathematics of students in grades 6–12 in nine dimensions: attitude towards success, confidence, anxiety, active involvement, beliefs about usefulness, perceiving mathematics as a male domain, and parents' and teachers' influence. Later,

Tapia and Marsh (Tapia & Marsh, 2002; Tapia, 1996) developed a questionnaire for high school and university students based on similar dimensions, aiming to make the instrument shorter and more relevant to older students. The instrument, called Attitudes toward Mathematics Inventory (ATMI) measures four dimensions: sense of security (later self-confidence), value, enjoyment, and motivation. Sense of security combines positive efficacy beliefs and feelings of anxiety, value denotes beliefs of usefulness of mathematics to oneself, enjoyment measures positive feelings and anticipations, and motivation describes willingness to choose and perform tasks involving mathematics. This shorter instrument forms the basis of measurements conducted in the current study.

3.2 Mathematical proving and attitudes

As attitudes have an important role in learning mathematics in general, one would assume that they also affect learning of proof. Stylianou et al. (2015) investigated students' views of the meaning of proof and how these views are related to students' attitudes, beliefs and experiences. They studied 535 students in six different universities, and found that high-performing students appreciated proofs more and had more positive beliefs about themselves as learners of proof than low-performing students. Furinghetti and Morselli (2007, 2009), on the other hand, investigated qualitatively how affect and cognition are intertwined in the proving processes of university students. Their studies imply that students' affects can act as driving forces in the proving process and either hinder or facilitate proof construction.

Iannone and Inglis (2010) designed a questionnaire that measures students' proof-related self-efficacy, that is, students' belief in their capability to produce proofs. In their study, Iannone and Inglis analysed the relationship of students' self-efficacy and their performance in proof tasks. They administered the questionnaire to 76 first-year students in a UK university, and found a positive correlation between students' self-efficacy at producing proofs and their actual proof production performance. Also, Viholainen et al. (2019) have investigated mathematics students' self-efficacy beliefs about proof. Their study comprised 29 Finnish and Swedish mathematics undergraduate students. Their results imply that students experienced high motivation towards proofs. At the same time, students doubted their own skills concerning proofs. Common reasons behind their low self-efficacy were problems in regulating the proving process and a fear of making mistakes. Selden and Selden (2013) have hypothesised that self-efficacy supports persistence in the process of constructing a proof, and the successes the students experience in writing proofs lead to a higher self-efficacy.

3.3 Background for attitude variables used in this study

The current study measured attitudes with a new instrument developed by combining and modifying two existing instruments (ATMI; Tapia, 1996; PSE; Iannone & Inglis, 2010). In this subsection, we characterise the four attitude variables measured by the new instrument and relate them to existing literature and related concepts. These variables are Self-efficacy, Anxiety, Appreciation and Motivation.

Self-efficacy is a student's belief in their capability to prove or learn proofs. The general concept of self-efficacy was studied extensively by Bandura (1977). In the Fennema–Sherman instrument, the similar construct was named “confidence in learning mathematics” (Fennema & Sherman, 1976). Self-efficacy is an accurate predictor of

academic achievement (e.g., Bandura & Schunk, 1981; Bartimote-Aufflick et al., 2016; Pajares & Graham, 1999; Zimmerman, 2000). It is linked to choice of activities, level of effort, persistence, emotional reactions and mathematical achievement (Bandura & Schunk, 1981; Pajares & Graham, 1999; Zimmerman, 2000). There are also implications that mathematics self-efficacy affects students' selection of science-based majors in college (Hackett & Betz, 1989). In secondary school, Pajares and Kranzler (1995) found that self-efficacy influences performance to a degree similar to general mental ability. Also, Arens et al. (2022) have reported positive reciprocal relations between math self-efficacy and math test scores. Regarding mathematical proving, Iannone and Inglis (2010) have shown that students' proof-related self-efficacy and performance correlate. Several studies have found gender differences in mathematics self-efficacy. In many studies, female students have been shown to have a lower self-efficacy than men (Betz & Hackett, 1983; Frost et al., 1994; Leder, 1995; Lahdenperä, 2018, Arens et al., 2022). According to Bandura's (1997) hypothesis, the development of self-efficacy is affected by a person's previous attainments, observing others, social persuasions, and emotional and physiological states.

Anxiety towards proving is a feeling of tension that interferes with proving activities. It can be compared to mathematics anxiety (for mathematics anxiety, see Richardson & Suinn, 1972). Numerous studies have shown a negative correlation between mathematics anxiety and student performance in all levels of education (e.g., Ma & Kishor, 1997; Ho et al., 2000; Miller & Bichsel, 2004). However, it is not clear what the causal direction between mathematics anxiety and performance is (e.g., Carey et al., 2016). Mathematics anxiety is closely linked to self-efficacy: people with low self-efficacy tend to experience high anxiety (Hoffman, 2010; Jain & Dowson, 2009; Pajares & Kranzler, 1995). A gender difference with respect to mathematics anxiety has been detected, and women are more likely to have mathematics anxiety than men (Devine et al., 2012; Hembree, 1990; Wigfield & Meece, 1988).

Appreciation of proof describes a student's view that proving and learning about proofs either has some utility value, such as developing reasoning skills, or is intrinsically valuable. Appreciation is closely related to usefulness in the Fennema–Sherman instrument (Fennema & Sherman, 1976). It also plays an important role in the expectancy–value theory of achievement motivation (Eccles et al., 1983). Using this framework, Greene et al. (1999) found that the combination of intrinsic and utility value predicted achievement in high school mathematics, both directly and through task-specific goal setting. Studies on gender differences in value beliefs about mathematics have yielded inconsistent findings (Gaspard et al., 2015). Regarding proving, Stylianou et al. (2015) found that university students appreciated the role of proof in mathematics, with high-performing students considering proofs more important than low-performing students.

Motivation refers in this study to a mostly emotional disposition. It describes the conscious desire to engage with proving-related tasks and challenges, based on enjoyment and interest. It relates therefore mostly to intrinsic motivation (e.g., Deci & Ryan, 1985), stemming from internal feelings of enjoyment, but also to effectance motivation (White, 1959), stemming from the desire to succeed and overcome challenges. Motivation has been found to correlate with achievement in mathematics, and the effect is bidirectional: perceptions of success increase motivation (Middleton & Spanias, 1999) and motivation predicts achievement (e.g., Murayama et al., 2013; Singh et al., 2002). Middleton and Spanias (1999) pointed out in their literature review that motivation towards mathematics forms around the middle school years and is relatively stable, but can be affected by careful instructional design. They also mention a well-known gender gap, with girls being less motivated to study mathematics.

4 Design and methods

4.1 Context

The study took place in the undergraduate course Introduction to University Mathematics in a Finnish research-intensive university. In Finnish universities, students declare a specific major when they enter the university and focus on that subject from the beginning or their studies. They also choose one or more minor subjects. The students of the course Introduction to University mathematics had mathematics as a major or minor subject. Common majors besides mathematics were computer science, economics, statistics and education. The course was a first-year course and typically the first university mathematics course for students. It lasted for one semester and was worth 5 ECTS (European Credit Transfer and Accumulation System) credits. The aim of the course was to support the transition from secondary to tertiary mathematics education. The main topics of the course included sets, functions, logic and proving. Proving was the main theme of the course, and students practised different kinds of proving strategies such as proving implications and equivalence, as well as indirect proofs and mathematical induction. Many of the students in the course studied other proof-based mathematics courses along with this course. The scale of the course was large: there were 600 students who showed some activity during the course (i.e., completed at least one task during the course).

The learning environment of the focus course was student-centred. It was implemented with the Extreme Apprenticeship method (Rämö et al., 2019, 2021) which is a form of inquiry-based mathematics education (Artigue & Blomhøj, 2013; Laursen & Rasmussen, 2019). Students started studying new topics by reading the course material and completing introductory, computerised tasks that gave them instant feedback. These tasks were based on the idea of proof frameworks by Selden et al. (2018), which help understanding the relationship between the logical structure of the statement to be proved and the structure of the proof. After the introductory tasks, the students deepened their understanding by completing more demanding pen-and-paper tasks.

Every other week one of the pen-and-paper tasks was assessed by the teaching team of the course. The teaching team comprised a responsible teacher and tutors who were undergraduate or graduate students. The team had weekly meetings in which mathematical topics and pedagogy were discussed. They gave feedback to the students on their solutions to the tasks, and the students could rewrite and resubmit their solutions. The students were offered an open learning space where they could spend as much time as they wanted. In the learning space, they could work collaboratively and receive help from the teaching team in completing the tasks and reading the course material. There were also lectures which focused on motivating the topics to the students, building the big picture and linking together different concepts. The importance of sharing one's ideas, even though they might not be correct, was emphasised to the students.

At the end of the course, students completed a final project, which was a broad assignment that focused on proving and explaining one's mathematical ideas in good mathematical style. The students had three weeks to complete the project. They could work together and ask for help from the teaching team, but the final submission had to be their own. The project contained parts requiring explaining in one's own words and applying concepts and ideas with mathematics they had not previously seen, so that answers could not be directly copied from other students. Key parts concerned understanding a written proof and writing one's own proofs. The project was assessed by the teaching team, and it counted for 15% of

the course grade. The rest of the grade was determined by the weekly tasks completed by the students.

4.2 Participants

The participants in this study were those students in the course Introduction to University Mathematics who responded to digital surveys during the course and gave their informed consent for using their responses for research purposes. Responses from all students were used to develop the attitude survey instrument as described in the next subsection. There were 535 students (89% of course participants) who responded to the pre-course questionnaire, and 440 students (73% of course participants) who responded to the post-course questionnaire.

In the main analysis, the number of participants was reduced for several reasons. Firstly, only those participants who responded to both pre- and post-questionnaires were included, so that the change in attitude variables could be investigated (RQ1). Secondly, only those students who gave their gender as either female or male were included, as they were the only groups large enough to be analysed statistically for differences (RQ2).

Thirdly, only those participants who had taken the advanced syllabus in mathematics in the high school final examination (matriculation examination) and provided their grade in the questionnaire were included. This was done in order to use the high school final examination grade as a control variable. Finally, only those students who submitted the final project in the focus course were included, so that the effect of attitude variables on performance could be studied (RQ3). In the end, the number of participants in the main analysis was $N=267$ (120 female, 147 male; 45% of course participants).

4.3 Data collection and factor analysis

Data was collected from participants using digital self-report surveys at the beginning and at the end of the focus course, that is, in September and in December of 2019. The attitude data was collected using a questionnaire based on two existing Likert scale questionnaires: the Attitudes Toward Mathematics Inventory (ATMI; Tapia, 1996) and Proof Self-Efficacy (PSE; Iannone & Inglis, 2010). The ATMI instrument measures four attitude variables in the context of mathematics: “sense of security,” “value,” “motivation” and “enjoyment,” using 40 question items and a 5-point Likert scale. The PSE instrument measures self-efficacy in the context of proving, using 10 items and a 5-point scale.

Both original instruments were translated into Finnish by one of the authors, and the items in ATMI were reworded to refer to proving instead of mathematics. For example, “I believe studying mathematics helps me with problem solving in other areas” was changed to “I believe that studying proofs helps me with problem solving in other areas” (VAL10). Those items in ATMI that could not be transformed were discarded. All items were combined into one set of 46 questions and their order was randomised. The 5-point answer scale was retained, ranging from “I strongly disagree” to “I strongly agree.” The question set was delivered in the same form and in the same order in both the pre- and post-course surveys.

Since the questionnaire was built from two separate instruments, of which only one focuses on proving and neither had been used in the Finnish context before, we decided to perform a factor analysis to extract the attitude dimensions that the instrument was measuring and to estimate the validity of the items. The factor analysis was initially performed

on the pre-course survey using R software (R Core Team, 2020) and the “psych” package (Revelle, 2020). We assumed a continuous scale for the responses and used Pearson correlation coefficients for the factor analysis, as suggested by Rhemtulla et al. (2012). To extract the factors, the maximum likelihood method was used, and since the attitudes were assumed to have non-zero correlations with each other, an oblique promax rotation was chosen (Costello & Osborne, 2005).

To find the number of factors, a scree plot together with parallel analysis of the eigenvalues were used. After making sure the suggested factors were distinct and coherent and reflected a theoretically justifiable construct, a 4-factor solution was chosen. Then, the set of items was reduced as follows. Items with a loading smaller than 0.6 were removed, resulting in a simple structure with each item loading mainly onto one factor. Some items loaded onto factors that conflicted with their original thematic contexts. For example, an item originally reflecting “enjoyment” in ATMI loaded onto the same factor as items reflecting “sense of security” and “self-efficacy.” Upon inspection, all such items were considered poorly worded or translated, and therefore removed. Lastly, the items with smallest loadings were removed to keep only six items for each factor. The final factor loadings are given in Table 8 in the Appendix. To confirm the factor structure, the same factor analysis was conducted to the post-course questionnaire. All items that were kept in the instrument loaded onto the same factors as in the pre-course questionnaire, so the factor structure was considered reliable. Finally, average scores were computed for the four factors to be used in the main analyses.

The factors were named based on their naming in the original instruments, as well as an interpretation of the items. The first factor contained all of the PSE items and some of the ATMI “sense of security” items. This factor was named Self-efficacy. The second factor contained the remaining ATMI “sense of security” items. These items reflected a negative sentiment and were negatively coded in the original instrument, so the factor was named Anxiety. The third factor contained only ATMI “value” items, so this factor was named Appreciation. The final factor contained ATMI “enjoyment” and “motivation” items. Based on the wording of the items finally chosen for the factor, the name Motivation was chosen. The origin and example items for each factor are given in Table 1. The complete item sets are included in Table 9 in the Appendix.

Table 1 The four attitude variables: Self-efficacy, Anxiety, Appreciation, and Motivation, their sources and example items in the final instrument

Variable	Source	Example items (translated from Finnish)
Self-efficacy	ATMI “sense of security,” PSE (self-efficacy)	PSE4: I am good at writing mathematical proofs SEC13: I believe that I am good at proving-related tasks
Anxiety	ATMI “sense of security,” negatively coded	SEC12: Studying proofs makes me feel nervous SEC6: Proofs make me feel uncomfortable
Appreciation	ATMI “value”	VAL10: I believe that studying proofs helps me with problem solving problems in other areas VAL4: Studying proofs is important and valuable VAL6: Strong skills in proving can help me in working life
Motivation	ATMI “motivation,” “enjoyment”	MOT3: The challenge of proving appeals to me ENJ4: I like solving new proving-related problems

4.4 Main analyses

Main analyses were conducted with R software (R Core Team, 2020), with the help of the package “rstatix” (Kassambara, 2020). To measure the relationship between high school grade and the attitude variables at the beginning of the course, we performed four linear regression analyses with the four attitudes as outcome variables and high school grade as the predictor. Next, we checked that there was no significant difference in the high school grades between the two genders using a Wilcoxon signed-rank test. This allowed us to compare the attitude variables between genders directly with four separate Student’s *t*-tests. Normality of residuals was tested for each attitude variable with the Shapiro–Wilk test and by examining quantile–quantile plots. Good levels of normality were observed for motivation and self-efficacy. For appreciation and anxiety, normality was poorly attested, but still considered reasonable for using *t*-tests. The variances of attitude scores in the two groups were equal, as assessed by Levene’s test. For each outcome variable, there were individual measurements that could be considered outliers, but we estimated that they would not have a large effect on the results. The *t*-tests were done using the Welch approximation to calculate the effective degrees of freedom for all attitude variables. Effect size was measured with Cohen’s *d* (small 0.2, medium 0.5, large 0.8).

To measure the change in the attitude variables during the course, we performed four mixed-design analyses of variance (ANOVA) with gender as the between-subjects variable and time of measurement (beginning or end of course) as the within-subjects variable. The outcome variables were the four attitudes. For the end-of-course measurements, only the self-efficacy variable had properly normally distributed residuals. However, the distributions of the other three attitude variables were considered acceptable for ANOVA. Motivation was normally distributed at the beginning but not at the end of the course. The variances of the attitude scores within the two genders were found to be equal using Levene’s test. Also, the covariances were homogeneous, as per Box’s *M*-test. Sphericity was tested using Maunchly’s test and corrected for as needed as part of the ANOVA procedure. Effect sizes were measured with generalised eta squared (small 0.01, medium 0.06, large 0.14).

To examine the effect of the attitude variables on performance, we used a binary logistic regression analysis with achievement level in the final course project as the outcome variable. The predictor variables were gender, high school grade, and the attitude variables measured at the beginning of the course. We chose to use the pre-course variables as predictors because the post-course responses might have already been affected by the project work, thus hindering causal inferences. The outcome variable was created by dividing the possible scores (0, 5, 10 or 15) into two categories: low (0, 5, 10) and high (15). The high category corresponded to good understanding of the course material and a decent ability to write and understand mathematical proofs. The low category corresponded to less understanding and a poor or modest ability to deal with proofs. With this division, both high and low categories had a similar number of students.

Throughout the analyses, we used the traditional alpha level of 0.05 for statistical significance. However, as this was exploratory research, we wrote out all *p*-values and considered them potentially informing up to 0.1. For the same reason, we did not apply any statistical correction for multiple testing, in order not to overlook any promising

candidates for further research (Streiner & Norman, 2011). However, we strove to avoid data dredging and making strong claims about results that were close to the chosen alpha level, on either side.

5 Results

5.1 Descriptive statistics

Table 2 gives the means, standard deviations and Cronbach's alpha coefficients of the attitude variables, as well as Pearson's correlation coefficients between the variables. The alpha coefficient measures the internal consistency of the variable, or the interrelatedness of the items comprising the variable. Information for gender (female/male), high school final examination grade, and performance in the final project (low/high) is also included in the table.

Before the main analysis, we observed from the means alone that Self-efficacy seemed to have increased during the course (from 2.70 to 3.01) and Anxiety diminished (from 2.43 to 2.25). Appreciation remained high throughout the course (3.87 pre-course, 3.92 post-course). Motivation has increased marginally (from 3.01 to 3.09). Anxiety had the largest standard deviation in both surveys (0.93 pre-course, 0.92 post-course), and Motivation had a similar standard deviation in the post-course survey (0.92). Cronbach's alpha coefficients were all above 0.88, which indicates strong internal consistency of the attitude variables.

Correlations between the attitude variables were all non-zero with $p < 0.01$. The strongest positive correlations were between Motivation and Self-efficacy (0.57 pre-course, 0.66 post-course), and between Motivation and Appreciation (0.59 pre-course, 0.56 post-course). The strongest negative correlations were between Anxiety and Self-efficacy (-0.55 pre-course, -0.64 post-course), and between Anxiety and Motivation (-0.55 pre-course, -0.62 post-course).

Based on the correlation coefficients, high school final examination grade correlated with all other attitude variables apart from Appreciation, having positive correlation with Self-efficacy and Motivation, and negative correlation with Anxiety. Gender correlated with Self-efficacy and Motivation, with male students tending to have larger values for both attitudes. Performance in the final project correlated positively with high school final examination grade, as well as with the same attitude variables as the high school examination grade.

5.2 Attitudes towards mathematical proof at the beginning of the course

To study the effect of previous performance on the attitude variables, a linear relationship was confirmed by regressing the four attitudes on the high school final examination grades. The results are given in Table 3. We found that high school grades had a positive linear relationship with Motivation and Self-efficacy, and a negative relationship with Anxiety. The relationship with Appreciation was not significant ($p > 0.05$).

After confirming the linear relationship between high school grades and the attitude variables, a non-parametric two-samples Wilcoxon rank test was done to see if there were any

Table 2 Means, standard deviations, Cronbach's alpha coefficients, and Pearson's correlation coefficients of the study variables

Variable	<i>M</i>	<i>SD</i>	α	1	2	3	4	5	6	7	8	9	10
1. Eff (pre)	2.70	0.76	0.90										
2. Anx (pre)	2.43	0.93	0.92	-0.55**									
3. App (pre)	3.87	0.75	0.90	0.35**	-0.32**								
4. Mot (pre)	3.01	0.83	0.88	0.57**	-0.55**	0.59**							
5. Eff (post)	3.01	0.81	0.91	0.69**									
6. Anx (post)	2.25	0.92	0.93	0.70**				-0.64**					
7. App (post)	3.92	0.81	0.91	0.71**		0.71**		0.29**	-0.27**				
8. Mot (post)	3.09	0.92	0.92		0.74**	0.66**		0.66**	-0.62**	0.56**			
9. HS grade (range 1-7)	5.56	1.25	-	0.29**	-0.14*	-0.02	0.21**	0.32**	-0.21**	0.06	0.20**		
10. Gender (f/m)	-	-	-	0.13*	-0.11	0.11	0.17**	0.16**	-0.08	0.08	0.18**	-0.00	
11. Project (low/high)	-	-	-	0.19**	-0.17**	0.07	0.23**	0.28**	-0.27**	0.08	0.26**	0.25**	0.03

For the attitude variables, correlation is reported within the timing context (pre- or post-course) for different variables, and between the timing contexts for the same variable. Attitude variables range from 1 to 5. Abbreviations: Eff = Self-efficacy, Anx = Anxiety, App = Appreciation, Mot = Motivation, pre = pre-course survey, post = post-course survey, HS grade = High school final exam grade, Project = performance in the course's final project. For correlations, gender is coded 1 = female, 2 = male, final project is coded 1 = low result, 2 = high result

* $p < 0.05$. ** $p < 0.01$

Table 3 Summary of linear regressions for high school grades predicting attitudes towards proving

Attitude	Linear coefficient of high school grade			Standardised coefficient β	t	p
	b	SE	95% CI			
Self-efficacy	0.17***	0.036	[0.10, 0.25]	0.29	4.89	<0.001
Anxiety	-0.12*	0.045	[-0.20, -0.02]	-0.14	-2.37	0.02
Appreciation	-0.01	0.037	[-0.08, 0.06]	-0.02	-0.31	0.76
Motivation	0.14***	0.040	[0.06, 0.22]	0.21	3.43	<0.001

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

Table 4 Summary of t -test results for the effect of gender on attitudes towards proving

Attitude	Female		Male		t	p	Cohen's d
	M	SD	M	SD			
Motivation	2.86	0.83	3.14	0.82	-2.14*	0.03	-0.34
Appreciation	3.77	0.73	3.94	0.76	-1.85	0.07	-0.23
Self-efficacy	2.59	0.77	2.79	0.75	-2.76**	0.006	-0.27
Anxiety	2.54	0.90	2.33	0.96	1.82	0.07	0.22

The Welch correction was used to calculate the effective degrees of freedom for all t -tests

* $p < 0.05$. ** $p < 0.01$

differences in high school grades between genders. No differences were found ($p = 0.93$). This test was chosen, as the grades were not normally distributed within either gender, according to the Shapiro–Wilk normality test ($p < 0.001$ in both groups).

Four t -tests were then performed to compare the effect of gender on attitudes towards proving. As presented in Table 4, male students exhibited significantly higher Self-efficacy ($t(253.0) = -2.76$, $p = 0.006$, Welch correction) than female students. Male students also had significantly higher Motivation ($t(252.0) = -2.14$, $p = 0.034$, Welch correction) than female students. Differences in Appreciation and Anxiety were non-significant. Effect sizes were small.

5.3 Changes in attitudes during the course

To assess the change in the attitude variables between the beginning and the end of the course among both genders, a mixed-design ANOVA was used, with time as a within-subjects variable and gender as a between-subjects variable. Table 5 shows the means and standard deviations of the attitude variables at both time points broken down by gender.

During the course, Self-efficacy increased significantly (large effect) and Anxiety decreased significantly (medium effect), as summarised in Table 6. The increase in Motivation was borderline significant ($p = 0.051$, small effect), whereas the change in Appreciation

Table 5 Means and standard deviations of the attitude variables at the beginning and at the end of the course, broken down by gender

Attitude	Beginning				End			
	Female		Male		Female		Male	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Motivation	2.86	0.83	3.14	0.82	2.91	0.91	3.24	0.90
Appreciation	3.77	0.73	3.94	0.76	3.85	0.83	3.99	0.80
Self-efficacy	2.59	0.77	2.79	0.75	2.86	0.78	3.13	0.83
Anxiety	2.54	0.89	2.33	0.96	2.33	0.89	2.18	0.94

Table 6 Summary for the mixed-design ANOVA comparing the attitudes towards proving at the beginning and at the end of the course and between genders

Predictor	Motivation			Appreciation		
	<i>F</i> (1, 265)	<i>p</i>	η^2_g	<i>F</i> (1, 265)	<i>p</i>	η^2_g
Time	3.83	0.05	0.002	2.51	0.11	0.001
Gender	9.54**	0.002	0.030	2.98	0.09	0.010
Time × Gender	0.43	0.51	<0.001	0.20	0.66	<0.001
Predictor	Self-efficacy			Anxiety		
	<i>F</i> (1, 265)	<i>p</i>	η^2_g	<i>F</i> (1, 265)	<i>p</i>	η^2_g
Time	62.89***	<0.001	0.036	17.00***	<0.001	0.010
Gender	7.02**	0.01	0.022	2.86	0.09	0.009
Time × Gender	0.80	0.37	<0.001	0.45	0.50	<0.001

η^2_g = generalised eta squared

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

was non-significant. There was no significant interaction between time and gender, meaning that the attitude variables changed similarly for female and male students.

5.4 Achievement predicted by prior attitudes, high school grade and gender

To analyse to what extent 1) gender, 2) high school grade and 3) attitudes at the beginning of the course predicted achievement in the final project of the course, a logistic regression was conducted. In the logistic regression, the predicted outcome was defined as “high” achievement (i.e., score of 15 on a discrete scale of 0, 5, 10, 15) in the final project. Table 7 presents a summary of the results.

High school grade and Motivation were the only significant predictors of achievement, when measured holding all other variables constant. The odds of getting a high score on the final project were increased by a factor of 1.5 per unit increase in high school grade, and a factor of 1.6 per unit increase in the Motivation score. The effects of gender, Self-efficacy, Anxiety and Appreciation on achievement were non-significant. The average scores in the final project were 10.6 for women and 11.0 for men.

Table 7 Binary logistic regression predicting high achievement in the final project of the course in terms of gender, high school grade, and attitudes towards proving at the beginning of the course

Predictor	Coefficient	SE	95% CI	<i>z</i>	OR	<i>p</i>
Gender ^a	-0.02	0.20	[-0.41, 0.37]	-0.09	0.98	0.93
High school grade	0.41***	0.10	[0.23, 0.61]	4.32	1.51	<0.001
Motivation	0.47**	0.18	[0.13, 0.82]	2.65	1.60	0.008
Appreciation	-0.19	0.16	[-0.50, 0.12]	-1.19	0.83	0.24
Self-efficacy	0.13	0.17	[-0.21, 0.46]	0.75	1.14	0.46
Anxiety	-0.19	0.14	[-0.47, 0.09]	-1.32	0.83	0.19

z = standardised coefficient, OR = odds ratio

^a 0 = female, 1 = male

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

6 Discussion

Values and attitudes are important drivers of studying and learning. In this study, we focused on beginning university students' attitudes towards proving, as proof-based reasoning is generally recognised as a particular challenge in the transition to university mathematics. Our results suggest that some attitude variables may indeed have an effect on performance and that they may themselves be affected by teaching or the learning environment more generally. We also found that attitudes towards proving are similar to attitudes towards mathematics in that they show a clear gender difference at the time of entering university.

6.1 Defining attitudes through measurement

Instead of deriving the attitude variables used in this study from a particular theory of attitude, we defined them by combining and modifying existing attitude instruments that we considered useful for this study. The resulting four attitude variables: Self-efficacy, Anxiety, Appreciation and Motivation, are all related to well-known mathematics affects. Self-efficacy and Appreciation can be seen as cognitive or metacognitive dispositions, with Self-efficacy directed to oneself ("I believe I can") and Appreciation to the object of learning ("I think it is useful"). Anxiety and Motivation, on the other hand, are more emotional, with Anxiety corresponding to almost visceral reactions ("I feel repulsed") and Motivation indicating desire to seek out proving-related situations in the future ("I find it interesting").

The new instrument performed relatively well overall, with clear distinctions between the four attitude dimensions and good internal consistency of each dimension. Self-efficacy and Motivation were approximately normally distributed. Appreciation was cut off from the high end (ceiling effect), having so many positive responses, and Anxiety was somewhat cut off from the low end (floor effect), with a less clear peak. These shortcomings should be addressed if the instrument were to be developed further.

6.2 Attitudes were connected with prior performance and gender

At the beginning of the course, students varied in their Self-efficacy beliefs, Anxiety towards proving and Motivation. Their high school grades correlated positively with

Self-efficacy and Motivation, and negatively with Anxiety. So, even though Finnish school mathematics does not include a lot of proofs, high school performance is linked with students' attitudes towards proving.

Our results align with Bandura's (1997) hypothesis of self-efficacy being affected by a person's prior experiences. They also corroborate the findings by Stylianou et al. (2015) who showed that high-performing students were more likely to hold more positive beliefs of themselves as learners of proof than low-performing students. Furthermore, our results are in line with previous studies that have shown that performance in mathematics is connected with self-efficacy (Hackett & Betz, 1989; Pajares & Graham, 1999), anxiety (Ma & Kishor, 1997; Ho et al., 2000; Miller & Bichsel, 2004) and motivation towards mathematics (Middleton & Spanias, 1999; Murayama et al., 2013; Singh et al., 2002). On the other hand, Appreciation of proofs did not vary a lot in the sample. Our result differs from the study by Stylianou et al. (2015), in which high-performing students appreciated proof more than low-performing students.

When we measured gender differences in attitude variables at the beginning of the course, we found that male students reported significantly higher Self-efficacy and Motivation than female students. The gender difference in Self-efficacy is in line with previous studies which have reported women having lower self-efficacy in mathematics than men (Betz & Hackett, 1983; Frost et al., 1994; Leder, 1995; Lahdenperä, 2018; Arens et al., 2022). It seems that the same gender gap exists with regard to proving. Similarly, Motivation towards studying mathematics is known to decrease among girls during middle school (e.g., Middleton & Spanias, 1999), and there may be a similar reason behind our results on proving.

We could not detect a significant gender difference in Appreciation or Anxiety towards proving. Many studies have found such a gender difference in mathematics anxiety (Hembree, 1990; Wigfield & Meece, 1988; Devine et al., 2012). In our study, there was an absolute difference of 0.2 standard deviation units in both Appreciation and Anxiety at the beginning of the course, males having a higher value in Appreciation and lower in Anxiety. It may be that our instrument was simply not sensitive enough to capture these differences as statistically significant with the current sample size. However, it may also be that students had not had a chance to experience proving at this stage, and therefore female students have not yet formed a negative response to it.

6.3 Interplay between learning environment and attitudes

During the course, students' Self-efficacy increased and Anxiety decreased on average. Also Motivation and Appreciation increased, but these changes were not statistically significant. As we did not have a control group and many students were taking other proof-based courses at the same time, we cannot conclude that these effects were due to this particular course. However, as proving is an activity rarely considered outside mathematics and the focus course was designed as an introductory course on proving, we are willing to hypothesise that at least part of the changes in these attitude variables were due to the learning environment in this course.

In the context of the focus course, there are several possible reasons for the increase in Self-efficacy that are in line with Bandura's (1997) model, which identifies prior attainments, vicarious experience and social persuasions among sources of self-efficacy. Learning proof was scaffolded with proof frameworks and tasks of increasing difficulty. It may be that these allowed students to experience early successes, which

affected their self-efficacy positively (see Selden & Selden, 2013). The students received support and encouraging feedback from the teaching team in a collaborative learning space. They also gave written peer feedback that allowed them to see their peers' work and compare their own capabilities to those of other students. Finally, the lecturer emphasised that making mistakes is part of the process of proving. In addition to increasing Self-efficacy, the same elements of the learning environment may also have decreased students' Anxiety (see Supekar et al., 2015).

In the final project, a high level of Motivation in the beginning of the course predicted good performance, after controlling for prior skills (high school grade) and while keeping the other attitudes fixed. We did not find a similar effect with the other three attitude variables, although in many previous studies both self-efficacy (Pajares & Graham, 1999) and anxiety (e.g., Ma & Kishor, 1997; Ho et al., 2000; Miller & Bichsel, 2004) have correlated with achievement. This may result from the fact that we controlled for prior performance in our analysis. For example, it may be that the students with high Self-efficacy did well in the final project, but if they also had strong skills at the beginning of the course, this effect was partially diminished in the analysis. We also acknowledge the high correlation between our attitude variables which may raise a problem of multicollinearity.

We conclude that Motivation was unique in that it helped students to surpass their prior skill level in the final project. As they had a relatively long time to work on the project by themselves and with support from the teaching team, high Motivation would probably help students work longer and harder, thereby achieving better results.

6.4 Implications for research and practice

Our study has indicated that attitudes towards proving can have an additional effect on performance in certain kinds of proving tasks. Also, students' attitudes towards proving can improve during their studies. These results make attitudes worthwhile to study and also to consider when planning teaching and assessment. The instrument developed in this study provides a quantitative tool for studying students' attitudes, which can be used and developed further in future studies. We suggest developing this tool towards one based on a comprehensive theoretical grounding, such as that of Di Martino and Zan (2010).

We detected an increase in students' Self-efficacy and a decrease in their Anxiety. In light of earlier literature, we hypothesise that collaborative learning environments in which the students interact with their peers and teachers has a positive effect on students' attitudes. Also, making the learning environment safe for the students by, for example, discussing factors that may cause anxiety and encouraging students to share their unfinished ideas, likely supports a beneficial development of attitudes. In the future, qualitative studies such as interviews and observations could reveal more about the reasons behind the changes in students' attitudes.

Our results show that, like in many areas of mathematics, there is a gender gap in students' attitudes towards proving: female students suffer from more detrimental dispositions towards proving. Teachers should be aware of this inequity and work actively towards changing the systemic deficiencies causing it. Based on prior research on gender gaps in students' attitudes towards mathematics, student-centred learning environments (Lahdenperä, 2018; Laursen et al., 2014) can help reduce the difference between females and males. However, when applying student-centred methods attention needs to be paid to the ways in which oppressive cultural narratives about women in mathematics affect women's ability to participate in the classroom (see Reinholz et al., 2022).

Appendix: Factor loadings and item translations

Table 8 shows the factor loadings of each item in the pre- and post-course surveys. In the table, the item names reflect the origin of the item: PSE refers to the PSE instrument (measuring self-efficacy), the other names refer to ATMI: SEC refers to “sense of security,” VAL to “value,” ENJ to “enjoyment” and MOT to “motivation.” Table 9 shows the question items in the final factor analysis solution, translated from Finnish.

Table 8 Factor loadings in pre- and post-course surveys

Item	Pre-course loadings				Post-course loadings			
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
PSE4	0.88	0.03	-0.03	0.00	0.90	-0.05	-0.01	-0.03
SEC13	0.86	-0.03	0.00	-0.03	0.80	-0.02	-0.04	0.12
PSE5	0.77	0.05	0.07	-0.01	0.76	0.05	0.10	-0.05
SEC2	0.73	-0.15	-0.01	-0.08	0.73	-0.16	0.02	-0.06
PSE2	0.68	0.08	-0.01	0.06	0.56	0.12	-0.02	0.27
SEC11	0.73	-0.09	0.03	0.01	0.74	-0.11	-0.01	0.00
SEC12	0.03	0.92	-0.04	0.09	0.02	0.89	0.06	-0.03
SEC6	0.03	0.86	0.03	-0.07	0.10	0.89	0.01	-0.11
SEC3	-0.08	0.80	-0.01	0.07	-0.05	0.84	-0.05	0.09
SEC5	0.02	0.76	0.00	0.01	-0.08	0.76	-0.01	0.05
SEC4	-0.12	0.74	0.07	0.03	-0.09	0.80	-0.01	0.10
SEC1	0.10	0.68	-0.05	-0.20	0.04	0.64	0.03	-0.33
VAL10	0.06	0.03	0.91	-0.12	0.04	0.02	0.91	-0.07
VAL7	0.04	-0.01	0.83	-0.10	0.03	0.01	0.86	-0.07
VAL8	0.01	-0.01	0.78	0.04	-0.08	-0.07	0.75	0.15
VAL2	-0.07	-0.06	0.74	0.05	0.04	0.03	0.81	0.01
VAL4	-0.04	0.03	0.66	0.19	-0.01	0.00	0.60	0.21
VAL6	0.03	0.02	0.70	0.02	0.02	0.01	0.81	-0.10
MOT3	-0.19	-0.01	0.00	0.94	-0.04	-0.02	0.15	0.73
ENJ4	0.13	-0.02	-0.04	0.80	0.08	-0.05	0.04	0.77
ENJ6	-0.09	-0.01	0.20	0.73	-0.09	-0.01	0.23	0.75
MOT2	0.03	0.04	0.06	0.65	0.02	0.03	-0.06	0.83
ENJ5	0.16	0.00	-0.07	0.53	0.15	0.04	-0.10	0.69
ENJ1	0.25	-0.07	0.02	0.52	0.12	-0.10	-0.02	0.73

The items are grouped according to the factors on which they loaded most strongly. Loadings above 0.5 are emphasised with boldface

Table 9 Question items retained after factor analysis, translated from Finnish

Self-efficacy	
SEC2	I am able to do proofs without too much difficulty
SEC11	I learn proving easily
SEC13	I believe I am good at solving proof problems
PSE2	I can often come up with my own proofs for statements I see in a lecture or class
PSE4	I am good at writing mathematical proofs
PSE5	It is easy for me to come up with convincing and logically sound mathematical arguments
Anxiety	
SEC1	When I hear the word "proof," I get a feeling of dislike
SEC3	It makes me nervous to even think about having to do a proof-related problem
SEC4	I am always under a terrible strain when working on proofs
SEC5	My mind goes blank and I am unable to think clearly when working on proofs
SEC6	Proofs make me uncomfortable
SEC12	Studying proofs makes me feel nervous
Appreciation	
VAL2	Studying proving is useful regardless of my field of study
VAL4	Studying proving is important and valuable
VAL6	Strong proving skills could help me in my career
VAL7	Proving is important for people besides mathematicians
VAL8	I think studying challenging proofs is useful
VAL10	I believe studying proving will help me with problem solving in other fields
Motivation	
ENJ1	I really like proving
ENJ4	I like solving new proof-related problems
ENJ5	I would prefer to do a proof problem than a problem related to another topic
ENJ6	Studying proving is very interesting
MOT2	I am willing to do extra proof problems
MOT3	I find the challenge of proving fascinating

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital. The work was partly funded by the Teachers' Academy of the University of Helsinki.

Data availability Due to ethical concerns, supporting data cannot be made available.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

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

References

- Aiken, L. R., Jr. (1970). Attitudes toward mathematics. *Review of Educational Research*, 40(4), 551–596. <https://doi.org/10.3102/00346543040004551>
- Alcock, L., & Weber, K. (2010). Undergraduates' example use in proof construction: Purposes and effectiveness. *Investigations in Mathematics Learning*, 3(1), 1–22. <https://doi.org/10.1080/24727466.2010.11790298>
- Almeida, D. (2000). A survey of mathematics undergraduates' interaction with proof: Some implications for mathematics education. *International Journal of Mathematical Education in Science and Technology*, 31(6), 869–890.
- Arens, A. K., Frenzel, A. C., & Goetz, T. (2022). Self-concept and self-efficacy in math: Longitudinal interrelations and reciprocal linkages with achievement. *The Journal of Experimental Education*, 90(3), 615–633. <https://doi.org/10.1080/00220973.2020.1786347>
- Artigue, M., & Blomhøj, M. (2013). Conceptualizing inquiry-based education in mathematics. *ZDM-Mathematics Education*, 45(6), 797–810. <https://doi.org/10.1007/s11858-013-0506-6>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, 84, 191–215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Bandura, A., & Schunk, D. (1981). Cultivating competence, self-efficacy, and intrinsic interest. *Journal of Personality and Social Psychology*, 41(3), 586–598. <https://doi.org/10.1037/0022-3514.41.3.586>
- Bartimote-Aufflick, K., Bridgeman, A., Walker, R., Sharma, M., & Smith, L. (2016). The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education*, 41(11), 1918–1942. <https://doi.org/10.1080/03075079.2014.999319>
- Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior*, 23(3), 329–345. [https://doi.org/10.1016/0001-8791\(83\)90046-5](https://doi.org/10.1016/0001-8791(83)90046-5)
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, 6, 1987. <https://doi.org/10.3389/fpsyg.2015.01987>
- Conner, A., Edenfield, K., Gleason, B., & Ersoz, F. (2011). Impact of a content and methods course sequence on prospective secondary mathematics teachers' beliefs. *Journal of Mathematics Teacher Education*, 14(6), 483–504. <https://doi.org/10.1007/s10857-011-9186-8>
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10, 7. <https://doi.org/10.7275/jyj1-4868>
- DeBellis, V. A., & Goldin, G. A. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, 63(2), 131–147. <https://doi.org/10.1007/s10649-006-9026-4>
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, 8(1), 1–9. <https://doi.org/10.1186/1744-9081-8-33>
- Di Martino, P., & Zan, R. (2015). The construct of attitude in mathematics education. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education: Exploring a mosaic of relationships and interactions* (pp. 51–72). Springer. https://doi.org/10.1007/978-3-319-06808-4_3
- Di Martino, P., & Zan, R. (2010). 'Me and maths': Towards a definition of attitude grounded on students' narratives. *Journal of Mathematics Teacher Education*, 13(1), 27–48. <https://doi.org/10.1007/s10857-009-9134-z>
- Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: A bridge between beliefs and emotions. *ZDM-Mathematics Education*, 43(4), 471–482. <https://doi.org/10.1007/s11858-011-0309-6>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). W. H. Freeman.
- Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7(5), 324–326. <https://doi.org/10.5951/jresmetheduc.7.5.0324>
- Frost, L. A., Hyde, J. S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics related attitudes and affect: A meta-analytic synthesis. *International Journal of Educational Research*, 21, 373–385. [https://doi.org/10.1016/S0883-0355\(06\)80026-1](https://doi.org/10.1016/S0883-0355(06)80026-1)

- Furinghetti, F., & Morselli, F. (2007). For whom the frog jumps: The case of a good problem solver. *For the Learning of Mathematics*, 27(2), 22–27.
- Furinghetti, F., & Morselli, F. (2009). Every unsuccessful solver is unsuccessful in his/her own way: Affective and cognitive factors in proving. *Educational Studies in Mathematics*, 70(1), 71–90.
- Furinghetti, F., & Morselli, F. (2011). Beliefs and beyond: Hows and whys in the teaching of proof. *ZDM-Mathematics Education*, 43, 587–599. <https://doi.org/10.1007/s11858-011-0316-7>
- Gaspard, H., Dicke, A. L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, 107(3), 663. <https://doi.org/10.1037/edu0000003>
- Greene, B. A., DeBacker, T. K., Ravindran, B., & Krows, A. J. (1999). Goals, values, and beliefs as predictors of achievement and effort in high school mathematics classes. *Sex Roles*, 40(5), 421–458. <https://doi.org/10.1023/A:1018871610174>
- Hackett, G., & 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/10.5951/jresmetheduc.20.3.0261>
- Hanna, G., & Barbeau, E. (2010). Proofs as bearers of mathematical knowledge. *ZDM-Mathematics Education*, 40, 345–353. <https://doi.org/10.1007/s11858-008-0080-5>
- Hannula, M. S. (2002). Attitude towards mathematics: Emotions, expectations and values. *Educational Studies in Mathematics*, 49(1), 25–46. <https://doi.org/10.1023/A:1016048823497>
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21, 33–46. <https://doi.org/10.5951/jresmetheduc.21.1.0033>
- Ho, H. Z., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., ... & Wang, C. P. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for Research in Mathematics Education*, 31(3), 362–379. <https://doi.org/10.2307/749811>
- Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, 45(1), 62–101. <https://doi.org/10.2307/749811>
- Hoffman, B. (2010). “I think I can, but I’m afraid to try”: The role of self-efficacy beliefs and mathematics anxiety in mathematics problem-solving efficiency. *Learning and Individual Differences*, 20(3), 276–283. <https://doi.org/10.1016/j.lindif.2010.02.001>
- Iannone, P., & Inglis, M. (2010). Self efficacy and mathematical proof: Are undergraduate students good at assessing their own proof production ability. In S. Brown (Ed.), *Proceedings of the 13th annual Conference on Research in Undergraduate Mathematics Education*. Retrieved April 30, 2023, from <http://sigmaa.maa.org/rume/crume2010/Archive/Iannone%20&%20Inglis.pdf>
- Jain, S., & Dowson, M. (2009). Mathematics anxiety as a function of multidimensional self-regulation and self-efficacy. *Contemporary Educational Psychology*, 34(3), 240–249. <https://doi.org/10.1016/j.cedpsych.2009.05.004>
- Kassambara, A. (2020). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (Version 0.6.0). <https://CRAN.R-project.org/package=rstatix>
- Lahdenperä, J. (2018). Comparing male and female students' self-efficacy and self-regulation skills in two undergraduate mathematics course contexts. In V. Durand-Guerrier, R. Hochmuth, S. Goodchild, & N. M. Hogstad (Eds.), *Proceedings of INDRUM 2018: Second conference of the International Network for Didactic Research in University Mathematics* (pp. 346–355). (*INDRUM Proceedings*). University of Agder and INDRUM. Retrieved April 30, 2023, from <https://indrum2018.sciencesconf.org/data/Indrum2018Proceedings.pdf>
- Laursen, S. L., & Rasmussen, C. (2019). I on the prize: Inquiry approaches in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education*, 5(1), 129–146. <https://doi.org/10.1007/s40753-019-00085-6>
- Laursen, S. L., Hassi, M. L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education*, 45(4), 406–418. <https://doi.org/10.5951/jresmetheduc.45.4.0406>
- Leder, G. C. (1985). Measurement of attitude to mathematics. *For the learning of Mathematics*, 5(3), 18–34. <https://www.jstor.org/stable/40247789>
- Leder, G. C. (1987). Attitudes towards mathematics. *The monitoring of school mathematics*, 2, 261–277. <https://eric.ed.gov/?id=ED289709>
- Leder, G. C. (1995). Equity inside the mathematics classroom: Fact or artifact? In W. G. Secada, E. Fennema, & L. B. Adajian (Eds.), *New directions for equity in mathematics education* (pp. 209–224). Cambridge University Press.
- Lesseig, K., & Hine, G. (2022). Teaching mathematical proof at secondary school: An exploration of pre-service teachers' situative beliefs. *International Journal of Mathematical Education in Science and Technology*, 53(9), 2465–2481. <https://doi.org/10.1080/0020739X.2021.1895338>

- Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, 28(1), 26–47. <https://doi.org/10.2307/749662>
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. *Handbook of Research on Mathematics Teaching and Learning*, 1, 575–596.
- Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, 30(1), 65–88. <https://doi.org/10.2307/749630>
- Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences*, 37(3), 591–606. <https://doi.org/10.1016/j.paid.2003.09.029>
- Mingus, T. T., & Grassl, R. M. (1999). Preservice teacher beliefs about proofs. *School Science and Mathematics*, 99(8), 438–444. <https://doi.org/10.1111/j.1949-8594.1999.tb17506.x>
- Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3), 249–266. <https://doi.org/10.1007/BF01273731>
- Murayama, K., Pekrun, R., Lichtenfeld, S., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, 84(4), 1475–1490. <https://doi.org/10.1111/cdev.12036>
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs and mathematics performance. *Contemporary Educational Psychology*, 24(2), 124–139. <https://doi.org/10.1006/ceps.1998.0991>
- Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20(4), 426–443. <https://doi.org/10.1006/ceps.1998.0991>
- R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.2). R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rämö, J., Reinholz, D., Häsä, J., & Lahdenperä, J. (2019). Extreme apprenticeship: Instructional change as a gateway to systemic improvement. *Innovative Higher Education*, 44(5), 351–365. <https://doi.org/10.1007/s10755-019-9467-1>
- Rämö, J., Lahdenperä, J., & Häsä, J. (2021). The extreme apprenticeship method. *Primus*, 31(10), 1106–1120. <https://doi.org/10.1080/10511970.2020.1818332>
- Reinholz, D., Johnson, E., Andrews-Larson, C., Stone-Johnstone, A., Smith, J., Mullins, B., Fortune, N., Keene, K., & Shah, N. (2022). When active learning is inequitable: Women's participation predicts gender inequities in mathematical performance. *Journal for Research in Mathematics Education*, 53(3), 204–226. <https://doi.org/10.5951/jresmetheduc-2020-0143>
- Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research* (Version 2.0.7). Northwestern University, Evanston, Illinois, USA. <https://CRAN.R-project.org/package=psych>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19(6), 551–554. <https://doi.org/10.1037/h0033456>
- Schwarz, B., Leung, I. C., Buchholtz, N., Kaiser, G., Stillman, G., Brown, J., & Vale, C. (2008). Future teachers' professional knowledge on argumentation and proof: A case study from universities in three countries. *ZDM-Mathematics Education*, 40(5), 791–811. <https://doi.org/10.1007/s11858-008-0150-8>
- Selden, A. (2012). Transitions and proof and proving at tertiary level. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education: The 19th ICMI study* (pp. 391–420). Springer. https://doi.org/10.1007/978-94-007-2129-6_17
- Selden, J., & Selden, A. (2009). Understanding the proof construction process. In F. L. Lin, F. J. Hsieh, G. Hanna & M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 Conference: Proof and proving in mathematics education* (Vol. 2, pp. 196–201). National Taiwan Normal University.
- Selden, A., & Selden, J. (2013). The roles of behavioral schemas, persistence, and self-efficacy in proof construction. In B. Ubuz, C. Haser, & M. A. Mariotti (Eds.), *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (pp. 246–255). Middle East Technical University.
- Selden, A., Selden, J., & Benkhalti, A. (2018). Proof frameworks: A way to get started. *Primus*, 28(1), 31–45. <https://doi.org/10.1080/10511970.2017.1355858>

- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95*(6), 323–332. <https://doi.org/10.1080/00220670209596607>
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology, 66*(3), 531–568. <https://doi.org/10.1111/peps.12035>
- Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: Is there a resolution? *Chest, 140*(1), 16–18. <https://doi.org/10.1378/chest.11-0523>
- Stylianou, D. A., Blanton, M. L., & Rotou, O. (2015). Undergraduate students' understanding of proof: Relationships between proof conceptions, beliefs, and classroom experiences with learning proof. *International Journal of Research in Undergraduate Mathematics Education, 1*(1), 91–134. <https://doi.org/10.1007/s40753-015-0003-0>
- Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. *Journal of Neuroscience, 35*(36), 12574–12583. <https://doi.org/10.1523/JNEUROSCI.0786-15.2015>
- Tapia, M. (1996). The Attitudes toward Mathematics Instrument. Paper presented at the *Annual Meeting of the Mid-South Educational Research Association*. <https://eric.ed.gov/?id=ED404165>
- Tapia, M., & Marsh, G. E. (2002). Confirmatory factor analysis of the attitudes toward mathematics inventory. Paper presented at the *Annual Meeting of the Mid-South Educational Research Association*. <https://eric.ed.gov/?id=ED471301>
- Viholainen, A., Tossavainen, T., Viitala, H., & Johansson, M. (2019). University mathematics students' self-efficacy beliefs about proof and proving. *LUMAT: International Journal on Math, Science and Technology Education, 7*(1), 148–164. <https://doi.org/10.31129/LUMAT.7.1.406>
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*(5), 297–333. <https://doi.org/10.1037/h0040934>
- Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology, 80*(2), 210.
- Zimmerman, B. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91. <https://doi.org/10.1006/ceps.1999.1016>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.