Relations Between Mathematics Self-Efficacy and Anxiety Beliefs: When Multicollinearity Matters

Erik Bergqvist

To cite this article: Erik Bergqvist (09 Apr 2024): Relations Between Mathematics Self-Efficacy and Anxiety Beliefs: When Multicollinearity Matters, The Journal of Experimental Education, DOI: 10.1080/00220973.2024.2338545

To link to this article: https://doi.org/10.1080/00220973.2024.2338545

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 09 Apr 2024.

Submit your article to this journal

View related articles

View Crossmark data
Relations Between Mathematics Self-Efficacy and Anxiety Beliefs: When Multicollinearity Matters

Erik Bergqvist

Department of Health, Education and Technology, Luleå University of Technology, Luleå, Sweden

ABSTRACT
This article reports on an investigation among 156 upper secondary students who self-evaluated their general mathematics self-efficacy and anxiety beliefs. After minimizing the influence of extreme multicollinearity, an exploratory factor analysis revealed a clear four-factor pattern. Two of the factors were interpreted as self-concept-like factors; one representing student's overall confidence in mathematics labeled mathematics self-concept, and the other labeled generalized mathematics self-efficacy, concerned with student's future-oriented perception of their competence in mathematics. The other two factors are students' anxiety when they give wrong answers in mathematics class (in-class anxiety), and students' anxiety and stress when they must do mathematics homework (assignment anxiety). Although no gender differences were found in previous mathematics achievement, the analyses indicated that all latent factors differed between genders. However, further analyses of logistic mixed models with all latent factors and prior mathematics achievement as predictors revealed that reported in-class anxiety was the primary discriminating factor between male and female students.

KEYWORDS
Factor analysis; mathematics anxiety; mathematics self-concept; mathematics self-efficacy; multicollinearity

Introduction
Self-constructs such as self-efficacy, self-concept, and anxiety are important in understanding motivation since they are a part of the overall perception of self (Skaalvik & Skaalvik, 2004). For example, in social cognitive theory, self-efficacy beliefs are particularly useful in explaining achievement outcomes (Bandura, 1982). However, despite an extensive amount of research, the results are inconsistent. For instance, educational research has shown that mathematics self-efficacy, self-concept, and anxiety are related to mathematics achievement, but to varying degrees (Byrne & Shavelson, 1986; e.g. Marsh et al., 2005; Pajares & Graham, 1999; Xie et al., 2019).

The inconsistency in results could stem from various factors, including the use of different assessment tools such as specific tests and school grades to measure mathematics achievement. In accordance with Pajares (1996) recommendation, it is crucial to synchronize the self-efficacy assessments and outcome measures with the specific area or task being studied to enhance explanatory and predictive capabilities of self-efficacy beliefs. Thus, general self-efficacy beliefs tend to be more powerful in predicting overall achievements such as grades compared to specific test scores.
Another aspect contributing to the variability in findings is the operationalization and measurement of the constructs, given the challenge in differentiating self-efficacy beliefs at a broader level of specificity (Bong & Skaalvik, 2003; Klee et al., 2022; Lee et al., 2020; Marsh et al., 2019). For instance, when studying mathematics self-efficacy, the measures often resemble those used for assessing self-concept (see e.g. Tossavainen et al., 2021), leading to ambiguity. Despite sharing similarities in predicting motivation and emotions based on experiences, self-efficacy and self-concept are conceptually distinct (Bong & Skaalvik, 2003; Marsh et al., 2019; Skaalvik & Skaalvik, 2004; Zimmerman, 2000). However, understanding this difference becomes more intricate when examining self-efficacy beliefs at a broader level of specificity, as often seen in educational research studies (Bong & Skaalvik, 2003).

The lack of clarity between general self-efficacy and self-concept poses challenges in operationalization, potentially leading to data characterized by high levels of multicollinearity. Given the sensitivity of factor or regression analysis to high multicollinearity (Field et al., 2013; Rockwell, 1975; Tabachnick & Fidell, 2013), addressing these concerns is crucial. Therefore, conducting exploratory investigations of the underlying factor structure before analysis is essential, even when using previously validated scales. Accordingly, this study aims to examine the factor structure of general mathematics self-efficacy and anxiety beliefs among students aged 16–19, while mitigating the impacts of high multicollinearity among the variables.

**Definitions, and issues with operationalization**

The self-concept is the individual description of self and involves judgments of the self that are based on past performances and contexts and have a strong connection to the social environment (Bong & Skaalvik, 2003; Marsh et al., 2019; Skaalvik & Skaalvik, 2004; Zimmerman, 2000). The mathematics self-concept is a domain-specific and pivotal construct in the present study, which is defined as the degree of certainty a person has regarding their mathematical competences (Skaalvik & Skaalvik, 2004). For example, this is measured with an item such as “I believe I am a person who is good at mathematics.”

Self-efficacy beliefs, on the other hand, concern an individual’s beliefs in their ability to complete a specific task, and they concern future-oriented conceptions (Bong & Skaalvik, 2003; Marsh et al., 2019; Skaalvik & Skaalvik, 2004; Zimmerman, 2000). According to Bandura (1982), “perceived self-efficacy is concerned with judgments of how well one can execute courses of action required to deal with prospective situations” (p. 122). Accordingly, mathematics self-efficacy is defined as an individuals’ confidence in completing a mathematical task (Pajares, 1996).

Marsh et al. (2019) proposed that self-concept-like constructs such as generalized mathematics self-efficacy, mathematics self-concept, and mathematics outcome expectations, differentiate from self-efficacy-like constructs such as pure mathematics self-efficacy, and mathematics functional self-efficacy. The latter two constructs are typically measured using more descriptive items. To respond to such items, students typically not required to make social comparisons. For example, wordings such as “How confident are you to be able to …” or “I can do …,” easily coincide with the operationalization of mathematics self-efficacy that were originally proposed by, e.g. Albert Bandura and Frank Pajares.

In the present study, generalized mathematics self-efficacy refers to more general self-efficacy perceptions without any specific criteria for successful performance. For example, a generalized mathematics self-efficacy measure would include items such as “I believe I can perform well on mathematics test.” Because it is unclear what it means to perform well, the person must use a frame of reference to respond properly. Accordingly, the construct is self-concept-like (Marsh et al., 2019).

Instruments measuring self-concept-like constructs include wordings that force judgment based on some frame of reference (Marsh et al., 2019). For example, to properly respond to wordings such as “I believe I can do well on a mathematics test,” the assessment must be put into some
context that is based on the individual’s frame of reference and social comparisons (Bong & Skaalvik, 2003; Marsh et al., 2019; Pajares & Schunk, 2002).

The empirical evidence from Marsh et al.’s (2019) study indicates the inability to distinguish between generalized mathematics self-efficacy and mathematics self-concept. However, this discovery may potentially stem from the ambiguous operationalization. For instance, the items “I am convinced that I can even understand the most difficult contents in math” and “It is easy to understand things in math” were utilized to measure generalized mathematics self-efficacy and mathematics self-concept, respectively. This is noteworthy because, despite any potential distinction, it could be too subtle for respondents to interpret these items differently.

Furthermore, similar issues also arise concerning various operationalization of self-concept and anxiety (Klee et al., 2022). For instance, there exists a strong correlation between mathematics self-efficacy, mathematics self-concept, and mathematics anxiety (Pajares & Kranzler, 1995). According to Ashcraft (2002), mathematics anxiety is defined by the bodily responses associated with feelings of worry and helplessness that interfere with mathematical performances elicited in, e.g., mathematical problem-solving situations. For example, students with low mathematics self-beliefs often worry that their mathematics abilities are not sufficient for succeeding in mathematics. Thus, mathematics anxiety negatively affects mathematics performance, because worrying thoughts interfere with the cognitive processes that take place in working memory, which in turn negatively affects the processing of the task itself (Ashcraft, 2002).

In general, anxiety can be conceptualized either as related to personal traits or situations (states) that elicit anxiety (Zeidner & Matthews, 2005). However, the conceptual differences are not distinct because the characteristic features are usually temporal rather than situational (Pekrun, 2006; Zeidner & Matthews, 2005). For instance, the apprehension associated with mathematics tests (test anxiety) may be based on specific circumstances, such as apprehension prior to a particular examination, or a more frequent emotion associated with testing situations (Pekrun, 2006).

### Issues with multicollinearity

Self-constructs with comparable levels of specificity, such as general mathematics self-efficacy beliefs and mathematics self-concept, exhibit a strong relationship. This close relationship may lead to a significant degree of multicollinearity among the measured variables. This refers to a situation where the observed variables are highly correlated with each other, which indicate redundancy or overlapping information (Tabachnick & Fidell, 2013). Interpreting results in quantitative research becomes particularly intricate when very high (extreme) multicollinearity is present during a regression or factor analysis (Field et al., 2013; Tabachnick & Fidell, 2013).

Sufficient intercorrelation between variables is necessary for conducting analyses such as factor analysis. However, excessive intercorrelation among variables in data poses a problem (Field et al., 2013; Rockwell, 1975; Tabachnick & Fidell, 2013). For example, in factor analysis, it occurs when two or more of the very highly related variables measure the same underlying construct, leading to difficulties in distinguishing their unique contributions and compromising the interpretability of the extracted factors (e.g. Field et al., 2013). Even though a larger sample size may mitigate the impact of sampling variability and offer more reliable estimates, it does not eliminate the underlying relationships among variables (Tabachnick & Fidell, 2013). Thus, extreme multicollinearity may be present even in large-scale studies.

Several researchers, among them Marsh et al. (2004), have highlighted the implications of very highly correlated variables in data. Accordingly, research constructing models predicting achievement from domain-specific self-efficacy beliefs may require reassessment due to the potential risk of having drawn inaccurate conclusions stemming from extreme multicollinearity in the data. For instance, the very high correlations observed in studies (e.g. Marsh et al., 2019; Pajares & Miller, 1994; Pietsch et al., 2003) suggest that even validated scales assessing general self-efficacy beliefs
need thorough scrutiny to avoid potential errors that would arise from excessively related or redundant variables. This scrutiny is particularly crucial given the acknowledged challenges in operationalizing domain-specific self-efficacy beliefs.

Although it is widely recognized in quantitative research that very high multicollinearity among variables in factor analysis poses challenges to factor interpretation (Tabachnick & Fidell, 2013), educational researchers rarely address these potential concerns. This matter often goes unnoticed, possibly because it is primarily considered a concern in small-scale studies. Nevertheless, certain studies confront the challenge posed by high multicollinearity among variables and delve into the essential data preparation for factor analysis. For example, Bergqvist et al. (2020) highlighted the importance of conducting a thorough analysis of highly correlated variables to identify potential sources of extreme multicollinearity before proceeding with a factor analysis.

Frequently, when exploring the interplay between self-concept-like constructs, issues arising from very high correlations between them are often resolved by eliminating one from regression analyses (see e.g. Pajares & Miller, 1994). However, a more suitable approach exists. Conducting exploratory analysis on all observed variables and then carefully removing one of the highly intercorrelated observed variables based on both qualitative and quantitative judgments (as suggested by Bergqvist et al., 2020) enables a thorough examination of the interplay among the self-concept-like constructs.

Relationships, and linkages with mathematics achievement

A vast number of research studies have explored self-concept-like constructs and their interrelationship, as well as their correlation with achievement. For instance, in the 1990s, Marsh (1990) and Marsh and Yeung (1997) found support for a reciprocal relationship between domain-specific self-concept, e.g. mathematics self-concept and achievement. That is, prior self-concept affects later achievement and prior achievements effect later self-concepts (referred to as the reciprocal effects model). In fact, the support for self-concept being both the cause and the effect of past performance is substantial in the literature, including a quite recently conducted meta-analysis by Wu et al. (2021).

In more recent studies, Holenstein et al. (2022) reported a significant correlation between mathematics self-concept and generalized mathematics self-efficacy \((r = 0.75)\), and past mathematics achievement \((r = 0.64 \text{ and } r = 0.45, \text{ respectively})\), and Marsh et al. (2019) reported very high intercorrelations between generalized mathematics self-efficacy and mathematics self-concept \((r > 0.90)\) after 4 years of testing among 3350 secondary school students. Marsh et al. (2019) conceptually differentiated between beliefs of mathematics self-concept and generalized mathematics self-efficacy. The empirical evidence indicates an inability to differentiate these constructs. Yet, without evaluating the extent of multicollinearity among variables, it remains uncertain whether the findings were influenced by extreme multicollinearity.

Furthermore, very high correlations have been observed between mathematics self-concept and mathematics anxiety. For instance, a very high average correlation of \(-0.71\) between mathematics self-concept and mathematics anxiety has been reported in a meta-analysis (Hembree, 1990). Furthermore, in recent studies, Goetz et al. (2010) reported high correlations \((r = -0.68)\) between mathematics self-concept and mathematics anxiety among students in grade 8. The high correlations suggest a large overlap between the constructs. In this regard, one study is especially noteworthy. Pajares and Miller (1994) reported a very high correlation \((r = -0.87)\) between mathematics self-concept and mathematics anxiety. Therefore, Pajares and Miller chose to exclude mathematics anxiety from the regression analyses. The rationale behind this decision was that the very high correlation led to instability in the parameter estimates. In other words, the constructs
seemed empirically indistinguishable, due to extreme multicollinearity. Other similar studies have also reported very high correlations between self-constructs (see e.g. Pajares & Graham, 1999).

In one study, the presence of high multicollinearity had significant consequences for the result. Pietsch et al. (2003) initially reported that generalized mathematics self-efficacy and self-concept were highly related ($r = 0.93$) and that generalized mathematics self-efficacy was the best predictor of achievement. However, a subsequent re-analysis by Marsh et al. (2004) concluded that this interpretation is incorrect. They highlighted that generalized mathematics self-efficacy and mathematics self-concept exhibited a very strong correlation, yet with very wide confidence intervals. Consequently, there were no reasons to assert that either construct was the superior predictor of achievement.

The explanatory power of self-efficacy judgments in predicting outcomes such as achievement depends on how closely aligned, both contextually and temporally, the predictor (self-efficacy) is with the outcome measure (Bong & Skaalvik, 2003; Pajares, 1996). For instance, Bandura (2005) and Pajares (1996) originally proposed that mathematics self-efficacy should be assessed by asking students to rate their confidence in completing a given mathematics test item, which has shown to have a stronger relation to specific test scores than school grades.

The inherent ambiguity in the relationships among mathematics self-constructs and the level of specificity in their operationalizations seem to contribute to the contradictory evidence regarding gender differences (Alves et al., 2016; Devine et al., 2012; Morán-Soto & González-Peña, 2022; Pajares & Graham, 1999; Xie et al., 2019). For instance, Hyde (2005) conducted a comprehensive meta-review, which primarily revealed gender similarities. These findings led to the proposition of the gender similarity hypothesis, suggesting minimal gender differences in psychological constructs. However, it is worth noting that some studies have yielded contradicting results when examining gender differences in mathematics self-beliefs, indicating that further investigations are necessary to fully understand this specific domain.

For example, Pajares and Miller (1994) conducted a study involving 350 university students and discovered that female students exhibited higher levels of mathematics anxiety, but lower task-specific mathematics self-efficacy than male students. Similarly, Pajares and Kranzler (1995) conducted an examination of the effect of mathematics anxiety and task-specific mathematics self-efficacy on problem-solving performance and only found a statistically significant gender difference in mathematics anxiety.

Further, in more recent research (Devine et al., 2012; Morán-Soto & González-Peña, 2022; Xie et al., 2019), female students reported higher levels of mathematics anxiety than male students. However, some studies have only found gender similarities between beliefs of mathematics anxiety and general mathematics self-efficacy (Alves et al., 2016; Pajares & Graham, 1999). Furthermore, Goetz et al. (2013) reported gender similarities in students’ mathematics achievement during two studies. In both studies, female students reported lower competence beliefs (generalized mathematics self-efficacy and mathematics self-concept) than male students.

In summary, prior research indicates that whether a gender gap exists depends on the specific context of the measurement (Else-Quest et al., 2010). For example, while girls typically report lower mathematics self-beliefs compared to boys (Zander et al., 2020), this discrepancy may diminish or vanish when controlling for factors such as previous achievement (Pajares, 2005). Therefore, the presence of a gender gap appears to rely on the variables included in regression models, highlighting the importance of addressing operationalization issues as they can impact the ability to predict outcomes.

### Aim and research questions

Considering the challenges presented by operationalization regarding evaluating beliefs of general mathematics self-efficacy and anxiety, the main objective is to investigate the factor structure of
these beliefs among a sample of upper-secondary school students. This is achieved by initially mitigating the influence of extreme multicollinearity in the data. Furthermore, the study aims to explore the significance of these factors in comprehending gender disparities.

The research questions for this article are as follows: (1) What are the underlying factors of upper secondary school student’s perceived general mathematics self-efficacy and anxiety beliefs? (2) How do these factors differ between female and male students?

The research questions are answered by analyzing data from upper-secondary students (aged 16–19) at a municipal school in Sweden. Data have been collected through questionnaires consisting of a set of Likert-type scales and analyzed with statistical methods. The subsequent section is devoted the presentation of the method and the results of the analyses of the factors that influence students’ beliefs of general mathematics self-efficacy and anxiety.

Method

Participants and procedure

In this article, data are drawn from two distinct studies. The first data sample is drawn from a study that aimed to explore mathematics self-efficacy and anxiety belief variables, and the consequences of removing highly correlating variables before a factor analysis. That study comprised a total of 79 students (57 female and 22 male students), who were either 16 or 17 years old and taking their first or second mathematics course. Bergqvist et al. (2020) have reported on the findings from this study.

Based on the findings by Bergqvist et al., 2020, a total of 17 questionnaire items were selected for a study that aimed to investigate students’ mathematics self-beliefs while committed to solving open-ended problems in mathematics. The second data sample was drawn from that study, which comprised 77 students, all of whom were 18 or 19 years old. These students were in their final semester of upper secondary school and taking their last mathematics course. In contrast to the study reported by Bergqvist et al. (2020), these students attended a mathematics course that was more advanced, both theoretically and mathematically.

Both data samples are drawn from studies that took place at a municipal upper secondary school in Sweden. In Sweden, the upper secondary school comprises years 10 to year 12 and the ages of students vary between 15 and 19 years. In total, the data analyzed in this article comprises responses from 156 upper secondary students (86 female and 70 male). The participating students have completed a self-reported questionnaire administered online by the author. The data collection has been conducted following the ethical guidelines of the Swedish Research Council. Students’ participation in the studies was completely voluntary, and they provided written informed consent to participate.

Measures

The Mathematics Self-Efficacy and Anxiety Questionnaire (MSEAQ; May, 2009) was used as an instrument to measure students’ generalized mathematics self-efficacy and anxiety beliefs. The students were requested to complete the survey with respect to their beliefs regarding their current mathematics class. This questionnaire was utilized as it comprises a diversity of different types of items, many of which the wordings are self-concept-like, such as “I believe I am the kind of person who is good at mathematics”. May (2009) proposed a five-factor model explaining the underlying structure of these variables among college students. However, when Chan and Abdullah (2018) investigated the same variables among primary school students, they found a unidimensional factor structure.
The questionnaire consists of 29 items (MSEAQ; May, 2009) with two subscales: Mathematics self-efficacy (mse), and mathematics anxiety (anx). The original items were written in English but translated into Swedish. In order to ensure reliability of the translated questionnaire items, Bergqvist et al. (2020) reported fit indices from a confirmatory analysis, which indicated a reasonable fit to the five-factor model originally proposed by May (2009).

The same five-point Likert-type scale was used throughout the questionnaire, with the following response options: 1 = aldrig (never), 2 = sällan (seldom), 3 = ibland (sometimes), 4 = ofta (often), 5 = nästan jämnt (usually). In contrast to the original response options, “usually” was translated into “nästan jämnt,” which back translated (word for word) is “almost always.” Furthermore, the “no response” option was not included. The reason for this was that students were expected to be able to evaluate statements regarding their mathematics self-belief.

Bergqvist et al. (2020) demonstrated that, based on an inspection of the correlation matrix, it is possible to reduce the number of items before conducting an Exploratory Factor Analysis (EFA). For example, two redundant items were removed: “I get tense when I prepare for a mathematics test.” and “I worry that I will not be able to do well on mathematics tests.” This was done because it seemed that students had assessed these items in the same way as they did with the item “I get nervous when taking a mathematics test.” In addition, two items were dropped during preliminary EFA: “I feel stressed when listening to mathematics instructors in class” and “I get nervous when I have to use mathematics outside of school” since they had only a very weak loading on any of the factors. Based on the validity and redundancy concerns and the data gathered, the initial study (reported on by Bergqvist et al., 2020) yielded 17 questionnaire items. These 17 items (see Table 1) are analyzed in this study.

The mse-items and anx-items (in Table 1) were found to be internally consistent (α’s were 0.89, and 0.78, respectively). Furthermore, students were asked for their legal gender. In Sweden, an individual’s legal gender is not considered sensitive personal information. Moreover, students were asked to provide their latest grade in mathematics. The mathematics grade was used as a measure of their prior mathematics achievement. In Sweden, mathematics students are graded on a scale from F (for failed), E, D, C, B, up to A (for the highest grade).

**Statistical analyses**

An EFA was applied to examine the underlying dimensions of the mathematics self-efficacy and anxiety variables. Although a previously validated scale was employed to assess students’ general mathematics self-efficacy and anxiety in this study, an exploratory approach (EFA) was chosen over a confirmatory one. This decision was influenced by the fact that students’ beliefs were assessed at a general level of specificity. This broader assessment inherits the challenges related to operationalization and multicollinearity. Consequently, exploring the factor structure becomes more appropriate than simply confirming it.

As Likert-type scales represent ordinal variables with ranked-ordered data, there is ongoing debate regarding the suitability of conducting EFA based on Pearson correlations (Choi et al., 2010). Notably, Pearson correlation coefficients can be influenced by deviations from normality, such as the presence of outliers (Wilcox, 2012). In light of these considerations, polychoric correlations emerge as a more appropriate choice. Polychoric correlations estimate the correlation between the normally distributed continuous underlying variables (Choi et al., 2010) and ignores the fact that the observed ordinal variables themselves are neither metric nor continuous.

To strengthen the argument in favor of polychoric correlations, the ordinal variables were assessed using Henze-Zirkler’s multivariate normality test. This test was facilitated by the R-package MVN (Version 5.9; Korkmaz et al., 2014). Statistically, significant results would support the conclusion that polychoric correlations are more appropriate.
To verify sampling adequacy for EFA, the Kaiser–Meyer–Olkin (KMO) measure was calculated, and the correlation matrix was checked to identify variables with very high intercorrelations and variables with lots of low intercorrelations. However, there are no statistical means for identifying very high correlations, only rules of thumb. Since variables with lots of high intercorrelations \( r_{ij} > 0.60 \) can also be a source of extreme multicollinearity (Rockwell, 1975). Therefore, the determinant of the correlation matrix was used as a heuristic tool to identify problems with extreme multicollinearity. If the determinant was below 0.00001, the correlation matrix was analyzed to identify the highest intercorrelations (cf. Field et al., 2013). According to the severity of the multicollinearity, one or both of the highest intercorrelating variables were removed from EFA.

Furthermore, Bartlett’s test of sphericity was used to test if correlations overall were large enough for EFA. However, although test results were significant, variables that had lots of low intercorrelations \( |r| < 0.30 \) were removed from EFA, cf. Field et al. (2013). Furthermore, the assessment of factor model fit was based on the values of TLI and RMSEA fit indices, with the cutoff values 0.06 and 0.95 for RMSEA and TLI, respectively (Hu & Bentler, 1999). In addition, the normality of the factor residuals was assessed using the Shapiro-Wilks normality test.

Moreover, the evaluation of the factor model’s quality was carried out using the criteria set forth by Costello and Osborne (2005) and Thompson (2004). These criteria stipulate that a robust factor in the model should possess a minimum of four items with high loadings and communalities, and no cross-loading.

### Table 1. The mathematics self-efficacy and anxiety items with their Swedish translations.

<table>
<thead>
<tr>
<th>Mathematics self-efficacy items</th>
<th>Swedish translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mse1</td>
<td>Jag tror jag kan förstå matematikkursens innehåll</td>
</tr>
<tr>
<td>mse2</td>
<td>Jag tror jag kommer att bli klar med alla uppgifter i matematikkursen.</td>
</tr>
<tr>
<td>mse3</td>
<td>Jag tror det kommer att gå bra när jag behöver använda mina matematikkunskaper i mitt framtida yrke.</td>
</tr>
<tr>
<td>mse4</td>
<td>Jag tror jag kan få betyg A i matematikkursen.</td>
</tr>
<tr>
<td>mse5</td>
<td>Jag tror jag kan tänka som en matematiker.</td>
</tr>
<tr>
<td>mse6</td>
<td>Jag tror det ska gå bra att lära mig matematik.</td>
</tr>
<tr>
<td>mse7</td>
<td>Jag tror det ska gå på matematikprov.</td>
</tr>
<tr>
<td>mse8</td>
<td>Jag tror jag är en person som är bra på matematik.</td>
</tr>
<tr>
<td>mse9</td>
<td>Jag känner mig själv säker när jag gör matematikprov.</td>
</tr>
<tr>
<td>mse10</td>
<td>Jag tror jag kan använda matematiken i kursen.</td>
</tr>
<tr>
<td>mse11</td>
<td>Jag känner mig tillräckligt själv säker för att ställa frågor under lektionen.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mathematics anxiety items</th>
<th>Swedish translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>anx12</td>
<td>Jag oroar mig för att jag inte kommer att bli klar med alla uppgifter i matematikkursen.</td>
</tr>
<tr>
<td>anx13</td>
<td>Jag är rätt för att säga fel under lektionen.</td>
</tr>
<tr>
<td>anx14</td>
<td>Jag blir nervös när det ställs frågor under lektionen.</td>
</tr>
<tr>
<td>anx15</td>
<td>Jag oroar mig för att det inte kommer att gå bra när jag behöver använda mina matematikkunskaper i mitt framtida yrke.</td>
</tr>
<tr>
<td>anx16</td>
<td>Jag blir stressad av att jobba med matematikuppgifter hemma.</td>
</tr>
<tr>
<td>anx17</td>
<td>Jag blir nervös när jag gör matematikprov.</td>
</tr>
</tbody>
</table>
All subsequent statistical analyses utilizing the identified latent factors were conducted using the student regression factor scores. Furthermore, Henze–Zirkler’s and the Shapiro–Wilks normality test were used to assess the pre-assumptions of multivariate and univariate normality of the factor scores. Based on these results, appropriate robust statistical tests (Wilcox, 2012) were applied.

Previous research has shown that self-concept-like constructs exhibit strong interrelationships and a strong correlation with prior mathematics grades (e.g. Holenstein et al., 2022). Therefore, the latent factors were further analyzed using a series of linear mixed effect models. In order to control for any disparities between the student responses originating from the three distinct mathematics courses, the mathematics course level is set as a random effect in these models.

Gender differences were analyzed using MANOVA. The most effective approach to comprehend the data in its entirety is to follow up a multivariate analysis of variance (MANOVA) with both univariate analysis of variance (ANOVA) and a logistic regression model or a discrimination analysis (Field et al., 2013). In the present study, the logistic model was used since gender is a binary variable. Accordingly, the differences between gender and prior mathematics achievement and the latent factors were analyzed using a series of logistic mixed effects models, with the mathematics course level as a random effect.

For all statistical tests, a \( p \)-value \(< .05\) was considered statistically significant. All statistical calculations and tests were performed using the statistical software R (Version 4.3.1; R Core Team, 2022) and the R-packages `correlation` (Version 0.8.4; Makowski et al., 2020), `lme4` (Version 1.1.34; Bates et al., 2015), `mice` (Version 3.16.0; van Buuren & Groothuis-Oudshoorn, 2011), `MVN` (Version 5.9; Korkmaz et al., 2014), `nlme` (Version 3.1.162; Pinheiro & Bates, 2000), `performance` (Version 0.10.4; Lüdecke et al., 2021), `psych` (Version 2.3.6; Revelle, 2022), `robustlmm` (Version 3.2.0; Koller, 2016), `WRS` (Version 0.24; Wilcox & Schönbrodt, 2014), and `WRS2` (Version 1.1.4; Mair & Wilcox, 2020).

**Results**

In one of the student responses, “sometimes” is reported consistently. Consequently, as it may indicate a lack of interest from the respondent, this response is excluded from the subsequent analyses. The sample from study 1 and study 2 contains 6% and 8% of missing values, respectively. Since multiple imputations are a recommended strategy for dealing with missing data (Brown, 2015), missing values are imputed using multiple imputations with a multinomial logit model provided by the R-package `mice` (Version 3.16.0; van Buuren & Groothuis-Oudshoorn, 2011).

Furthermore, only one student reported a mathematics school grade of F (\( F = \) failed). This response is included in the EFA, but it is removed from subsequent analyses. After the removal, the number of responses distributed among mathematics grades are 39 (E), 31 (D), 44 (C), 26 (B), and 14 (A). The majority of the students (76) are enrolled in the advanced mathematics course, whereas 43 and 36 students are enrolled in one of the initial two (basic) mathematics courses, respectively.

The following sections present different sets of analyses. The first section reports on the preparation before EFA, i.e. the preliminary analysis of data. The results from the conducted EFA are followed by an evaluation of the latent factor score distributions and an analysis of the relative contributions from prior mathematics achievement and gender. In the last section of the results, the findings from the gender differences between the latent variables are presented.

**Preliminary analysis**

The Kaiser–Meyer–Olkin (KMO) measure is good (\( KMO = 0.89\)), thus in the “meritorious” range according to Kaiser (1974). The KMO values for all the individual items are well above 0.50, which is considered the acceptable minimum (Kaiser, 1974; Kaiser & Rice, 1974). However, the
determinant of the correlation matrix is 0.000008, which is less than the heuristic 0.00001, and thus indicates problems with extreme multicollinearity. Therefore, the correlation matrix is also inspected to find the highest intercorrelations. Henze-Zirkler’s multivariate normality test \( HZ = 1.004, p < .001 \) indicate that the ordinal variables are not multivariate normal, thereby indicating that polychoric correlations are more appropriate than Pearson’s correlations.

In the correlation matrix (Table 2), three variables are identified as having the highest intercorrelations (\(|r| \geq 0.75\)): mse4: “I believe I can get an ‘A’ when I am in a mathematics course”, mse7: “I believe I can do well on a mathematics test”, and mse9: “I feel confident when taking a mathematics test”. Both variables mse7 and mse9 (\( r = 0.78 \)) seem to assess students’ reported confidence while taking a mathematics test, which in turn is related to variable mse4 that assesses students’ confidence in getting the highest grade in the mathematics course (\( r = 0.77 \), and \( r = 0.76 \), respectively). Since these variables are highly correlated and the determinant shows problems with multicollinearity, variable mse7 is removed from EFA. After removal, the determinant value is 0.00003, which is acceptable.

Furthermore, Bartlett’s test of sphericity, \( \chi^2(136) = 1343, p < .001 \) is statistically significant which show that correlations overall are not too low. The variable mse2: “I believe I can complete all of the assignments in a mathematics course” has lots of low intercorrelations and is therefore also removed from EFA. The preliminary analysis indicates that the sample is adequate for EFA, but two problematic variables mse7 and mse2 should be removed before conducting the EFA.

**Exploratory factor structure**

Based on the parallel analysis (as depicted in Figure 1), four factors are retained. Furthermore, since the latent factors are expected to be highly correlated, the minimum residuals method with oblique rotation (promax) is applied to reveal the factor pattern.

EFA is conducted three times using the `fa()` function from the R-package *psych* (Version 2.3.6; Revelle, 2022), and each time a variable with cross-loading above 0.30 is removed. After the removal of variables anx15, mse9, and mse6, the final factor model explains 65% of the observed variance. Table 3 presents a summary of the final factor model with the factor loadings

<table>
<thead>
<tr>
<th>Table 2. Polyehoric correlation matrix.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mse1</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>mse2</td>
</tr>
<tr>
<td>mse3</td>
</tr>
<tr>
<td>mse4</td>
</tr>
<tr>
<td>mse5</td>
</tr>
<tr>
<td>mse6</td>
</tr>
<tr>
<td>mse7</td>
</tr>
<tr>
<td>mse8</td>
</tr>
<tr>
<td>mse9</td>
</tr>
<tr>
<td>mse10</td>
</tr>
<tr>
<td>mse11</td>
</tr>
<tr>
<td>anx12</td>
</tr>
<tr>
<td>anx13</td>
</tr>
<tr>
<td>anx14</td>
</tr>
<tr>
<td>anx15</td>
</tr>
<tr>
<td>anx16</td>
</tr>
<tr>
<td>anx17</td>
</tr>
</tbody>
</table>

Note: \( N = 155 \). Significance levels.

- \( p < .001 \)
- \( p < .01 \)
- \( p < .05 \) (two-tailed). Intercorrelations over 0.75 appear in bold.
regression coefficient) and the corresponding structure coefficient (correlation) between each variable and the factor, which is shown in the parentheses.

The factor model fit index values RMSEA and TLI are 0.057 and 0.96, respectively. Furthermore, the factor residuals are normally distributed, $W = 0.98$, $p = 0.27$ (Shapiro-Wilk test), and none of the factor residuals are greater than 5%, which altogether suggests a good factor model fit (Field et al., 2013). All four factors are internally consistent with Cronbach’s $\alpha$: 0.82, 0.78, 0.74, and 0.73.

**Table 3.** Summary of EFA results including factor intercorrelations.

<table>
<thead>
<tr>
<th>Questionnaire item/factor intercorrelations</th>
<th>MSC (18.7 %)</th>
<th>ICA (17 %)</th>
<th>GMS (15.9 %)</th>
<th>ASA (13.3 %)</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Self-Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe I can think like a mathematician.</td>
<td><strong>0.89</strong></td>
<td>0.06(-0.28)</td>
<td>-0.02(0.48)</td>
<td>0.06(-0.43)</td>
<td>0.67</td>
</tr>
<tr>
<td>I believe I am the kind of person who is good at mathematics.</td>
<td><strong>0.75</strong></td>
<td>-0.11(-0.47)</td>
<td>0.05(0.57)</td>
<td>-0.05(-0.58)</td>
<td>0.77</td>
</tr>
<tr>
<td>I believe I can get an ‘A’ when I am in a mathematics course.</td>
<td><strong>0.70</strong></td>
<td>0.07(-0.29)</td>
<td>0.35(0.70)</td>
<td>0.10(-0.45)</td>
<td>0.74</td>
</tr>
<tr>
<td>In-Class Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am afraid to give an incorrect answer during my mathematics class.</td>
<td>-0.10(-0.34)</td>
<td><strong>0.86</strong></td>
<td>0.10(-0.20)</td>
<td>-0.10(-0.27)</td>
<td>0.69</td>
</tr>
<tr>
<td>I get nervous when asking questions in class.</td>
<td>-0.06(-0.36)</td>
<td><strong>0.83</strong></td>
<td>0.20(-0.17)</td>
<td>0.13(0.41)</td>
<td>0.75</td>
</tr>
<tr>
<td>I feel confident enough to ask questions in my mathematics class.</td>
<td>-0.25(0.22)</td>
<td>-0.71(-0.69)</td>
<td>0.39(0.42)</td>
<td>0.10(-0.23)</td>
<td>0.57</td>
</tr>
<tr>
<td>Generalized Mathematics Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe I can understand the content in a mathematics course.</td>
<td>0.03(0.53)</td>
<td>0.07(-0.24)</td>
<td><strong>0.80</strong></td>
<td>-0.07(-0.45)</td>
<td>0.70</td>
</tr>
<tr>
<td>I believe I will be able to use mathematics in my future career when needed.</td>
<td>0.07(0.54)</td>
<td>0.04(-0.26)</td>
<td><strong>0.63</strong></td>
<td>-0.17(-0.50)</td>
<td>0.57</td>
</tr>
<tr>
<td>I believe I can do the mathematics in a mathematics course.</td>
<td>0.30(0.58)</td>
<td>-0.09(-0.34)</td>
<td><strong>0.46</strong></td>
<td>0.06(-0.38)</td>
<td>0.47</td>
</tr>
<tr>
<td>Assignment Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working on mathematics homework is stressful for me.</td>
<td>0.18(-0.39)</td>
<td>-0.11(0.24)</td>
<td>-0.10(-0.41)</td>
<td><strong>0.95</strong></td>
<td>0.73</td>
</tr>
<tr>
<td>I worry that I will not be able to complete every assignment in a mathematics course.</td>
<td>-0.14(-0.58)</td>
<td>-0.03(0.32)</td>
<td>-0.18(-0.53)</td>
<td><strong>0.57</strong></td>
<td>0.59</td>
</tr>
<tr>
<td>I get nervous when taking a mathematics test.</td>
<td>-0.13(-0.50)</td>
<td>0.30(0.53)</td>
<td>0.09(-0.33)</td>
<td><strong>0.50</strong></td>
<td>0.53</td>
</tr>
<tr>
<td>MSC: Mathematics Self-Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICA: In-Class Anxiety</td>
<td>-0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMS: Generalized Mathematics Self-efficacy</td>
<td>0.61</td>
<td>-0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA: Assignment Anxiety</td>
<td>-0.60</td>
<td>0.42</td>
<td></td>
<td>-0.48</td>
<td></td>
</tr>
</tbody>
</table>

Note: $N = 155$. $h^2$ = communalities. Factor loading over 0.40 appear in bold. Structure coefficients appear in parentheses.
The factor interpretations

The factor intercorrelation coefficients (as depicted in Table 3) demonstrate that the relationship between MSC factor, GMS, and ASA factor is strong (about 36% shared variance).

The items that load on the MSC factor explain 18.7% of the variance and represents students’ overall confidence in being good at mathematics. The factor is labeled mathematics self-concept (Bong and Skaalvik, 2003; Marsh et al., 2019; Skaalvik & Skaalvik, 2004).

The ICA factor (17%) reflects students’ worries about giving incorrect answers during class and is therefore labeled in-class anxiety. The ICA factor has a significant cross-loadings on the GMS factor, which might indicate a problem with that item. However, that item is kept on the ICA factor since the loading on the corresponding factor is very high and fits very well in the description of the factor.

The items loading on the GMS factor (15.9%) concern students’ beliefs in being able to understand and use mathematics. These items reflect future-oriented conceptions of their competence, that are typically utilized for evaluating mathematics self-efficacy beliefs (Bong and Skaalvik, 2003; Marsh et al., 2019). However, as these items do not specify any criteria for performance, it also forces students to evaluate their beliefs based on some frame of reference. Hence, this factor is interpreted as a construct that resembles mathematics self-concept rather than mathematics self-efficacy and consequently referring to generalized mathematics self-efficacy (Marsh et al., 2019; Skaalvik & Skaalvik, 2004).

Finally, the fourth factor (13.3%) is labeled as assignment anxiety (ASA). It describes students’ worries about having to do mathematics homework, and not being able to complete every assignment in the course. The item “I get nervous when taking a mathematics test” also loaded on this factor, which indicate that this factor is related to students’ test anxiety. This finding is not surprising since the stress in having to do mathematics homework may also reflect students’ worries that their abilities are not sufficient to succeed in a testing situation (Zeidner & Matthews, 2005). Thus, this factor reflects students’ worries about being confronted with situations when their mathematics competence is being evaluated. One such situation is students’ stressfulness in doing homework, being reported by the item “Working on mathematics homework is stressful for me.”

Evaluation of factor scores

The correlation of regression scores with for all four factors are greater than 0.91, which indicate that the factor scores are adequate (Grice, 2001). Furthermore, Figure 2 presents differences in the regression factor score distributions for the latent variables. In general, the factor scores appear to be fairly balanced. Nonetheless, mathematics self-concept (MSC), assignment anxiety

![Figure 2. Factor score distributions with imposed jittered scatterplots.](image-url)
(ASA), and in-class anxiety (ICA) have right-skewed distributions. However, using the estimated interquartile range for each variable indicates that there are no extreme values (outliers). Still, within female students, possible outliers are identified for mathematics self-concept (MSC) and generalized mathematics self-efficacy (GMS).

Further, the Henze-Zirkler’s tests for female students \((HZ = 1.02, \ p = .026)\), and male students \((HZ = 1.09, \ p = .006)\) are statistically significant which indicates that data not being multivariate normal within gender. Furthermore, according to the Shapiro-Wilk test, the overall factor score distributions for GMS show evidence of normality \((\text{Shapiro-Wilk test}), W = 0.99, p = 0.41\). However, for factor score distributions MSC \((W = 0.96, p < .001)\), ICA \((W = 0.96, p < .001)\) and ASA \((W = 0.98, p = .020)\), the test results are statistically significant, which indicates non-normal distributions.

To summarize, the interpretation of the distributions of factor scores depicted in Figure 2 and the results from Henze-Zirkler’s tests indicate that the factor scores within gender are not multivariate normally distributed. Furthermore, only the GMS factor scores show evidence of being normally distributed.

**Latent factor predictors**

First, it is reasonable to assess the need for a mixed effects model using deviance \((-2 \times \text{log-likelihood})\) statistics \((\text{Field et al., 2013})\). Because if there is no evidence of variance across mathematics course levels, a regression analysis would be more straightforward. Therefore, the chi-square statistics are estimated for the difference in deviance between a generalized least squares model that only includes the intercept, and the corresponding model with random intercept only. Both models are fit using the \(\text{gls()}\) and \(\text{lme()}\) function, respectively, provided by the R-package \(\text{nlme}\) \((\text{Version 3.1.162}; \text{Pinheiro & Bates, 2000})\).

For the ICA factor, there is clearly no evidence of variance across mathematics course levels, \(\chi^2(1) = 0.002, p = 0.96\). However, the statistically significant chi-square tests for the remaining models indicate the need for mixed effect models: For MSC factor, \(\chi^2(1) = 15.21, p < .001\), GMS factor, \(\chi^2(1) = 4.84, p = 0.028\), and ASA factor, \(\chi^2(1) = 6.81, p < .01\).

Table 4 shows the results from a series of linear mixed effect models predicting the four factors. The mathematics course level is set as a random effect. The predictor variables are gender and prior mathematics achievement. Evidence suggests that the GMS factor meets the assumption of normally distributed factor scores (as depicted in Figure 2). Therefore, a conventional linear effect model \((\text{provided by R-package \text{nlme}4}; \text{Bates et al., 2015})\) is applied with mathematics course level as a random effect, including restricted maximum likelihood (REML) as parameter estimate \((\text{Bates et al., 2015})\). The model fit is statistically significant, \(\chi^2(5) = 62.79, p < .001\).

However, despite the absence of outliers, a robust linear effects model is applied due to the presence of skewness for the MSC, ICA, and ASA factor score distributions. The parameter estimates for this robust model are provided by the R-package \textit{robustlmm} \((\text{Version 3.2.0}; \text{Koller, 2016})\). The log-likelihood statistics are unavailable for this type of model \((\text{Koller, 2016})\) and accordingly cannot been utilized for model fit assessment, and estimation of profile confidence intervals. Instead, these robust models are evaluated by examining the residuals, and the confidence intervals are estimated using the Wald \(r\)-distributions. According to the quantile-comparison plots \((\text{see Table 4})\), where the norm quantiles are plotted against studentized residuals, the factor residuals for all the models are approximately normally distributed, which indicates a good model fit \((\text{Field et al., 2013}; \text{Tabachnick & Fidell, 2013})\).

In Table 4, Nakagawa’s marginal and conditional \(r\)-squared values for mixed models are provided by the R-package \textit{performance} \((\text{Version 0.10.4}; \text{Lüdecke et al., 2021})\). The conditional \(r\)-squared value encompasses both random and fixed effects, whereas the marginal \(r\)-squared value, as indicated in parentheses, solely considers the variance of fixed effects \((\text{Lüdecke et al., 2021})\).
Table 4. Summary of the linear mixed models and residual analyses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>MSC</th>
<th></th>
<th></th>
<th>ICA</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>GMS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>ASA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ [95% CI]</td>
<td>t</td>
<td>$\beta$ [95% CI]</td>
<td>t</td>
<td>$\beta$ [95% CI]</td>
<td>t</td>
<td>$\beta$ [95% CI]</td>
<td>t</td>
<td>$\beta$ [95% CI]</td>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$-0.38 [-0.93, 0.16]$</td>
<td>$-1.39$</td>
<td>$0.30 [0.02, 0.58]$</td>
<td>$2.07$</td>
<td>$-0.16 [-0.50, 0.21]$</td>
<td>$-0.88$</td>
<td>$0.36 [-0.01, 0.72]$</td>
<td>$1.91$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade E</td>
<td>$-0.48 [-0.80, -0.16]$</td>
<td>$-2.92$</td>
<td>$0.41 [0.01, 0.80]$</td>
<td>$2.00$</td>
<td>$-0.59 [-0.92, -0.24]$</td>
<td>$-3.41$</td>
<td>$0.27 [-0.12, 0.66]$</td>
<td>$1.36$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade B</td>
<td>$0.66 [0.31, 1.01]$</td>
<td>$3.67$</td>
<td>$0.00 [-0.43, 0.43]$</td>
<td>$0.00$</td>
<td>$-0.31 [-0.67, 0.01]$</td>
<td>$-1.62$</td>
<td>$0.06 [-0.35, 0.52]$</td>
<td>$0.38$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade A</td>
<td>$1.48 [1.02, 1.94]$</td>
<td>$6.31$</td>
<td>$-0.80 [-1.35, -0.25]$</td>
<td>$-2.86$</td>
<td>$1.18 [0.71, 1.71]$</td>
<td>$4.75$</td>
<td>$-0.84 [-1.39, -0.28]$</td>
<td>$-2.95$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>$0.49 [0.24, 0.73]$</td>
<td>$3.89$</td>
<td>$-0.75 [-1.05, -0.45]$</td>
<td>$-4.95$</td>
<td>$0.28 [0.03, 0.56]$</td>
<td>$2.13$</td>
<td>$-0.63 [-0.93, -0.34]$</td>
<td>$-4.19$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$ (Nakagawa)</td>
<td>0.53(0.36)</td>
<td></td>
<td>0.23(0.23)</td>
<td></td>
<td>0.36(0.31)</td>
<td></td>
<td>0.25(0.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $N=154$. Dummy coding: Gender (0 = Female, 1 = Male), prior math achievement (reference group grade C). The course level is a random effect. The norm quantiles are plotted against studentized residuals with 95% confidence interval. Statistically significant estimates appear in bold.
For instance, the model predicting the ICA factor has the same value (0.23) for both the marginal and the conditional $r$-squared value. This is yet another indication of consistency across mathematics course levels. In contrast to e.g. the model predicting MSC, which has a significant random effect.

Moreover, for the GMS model, the 95% confidence intervals are estimated using profile likelihood confidence intervals, whereas for the remaining models, it is derived from the Wald $t$-distribution. In all the models (in Table 4), grade C is used as a reference group. According to the summary in Table 4, both GMS and GMS are positively predicted by the highest grades (B and A) compared to grade C (reference group). For instance, grade A is especially a strong predictor of MSC and GMS ($\beta = 1.48$ and 1.18, respectively). Therefore, the result indicates that high past mathematics achievement has a larger effect on the MSC factor.

Further, low prior mathematics achievement (grade E) negatively predicts GMS ($\beta = -0.59$) more strongly than MSC ($\beta = -0.48$). Negative effects are also found for high past mathematics achievement on ASA ($\beta = -0.84$ and $-0.44$, respectively. However, since some CIs almost contain zero, they should be interpreted with caution. For instance, the corresponding CI exhibits an end value that is very close to zero, even though grade E has a direct positive effect on ICA, whereas grade B has a negative effect on ASA. Furthermore, all latent factors are predicted by gender when prior mathematics achievement is controlled for, even though the GMS factor shows indications of having the weakest effect ($\beta = 0.28$).

**Analyses of gender differences**

The distribution of factor scores within gender violates normality assumptions. For instance, several possible outliers are detected, which could bias the results when using statistical tests that are based on the mean (Wilcox, 2012). Consequently, following the recommendations by Wilcox (2012), gender differences are analyzed using MANOVA based on 20% trimmed means. The results (provided by R-package WRS2; Wilcox & Schönbrodt, 2014) show a statistically significant effect of gender on the latent variables, and prior mathematics achievement $F = 46.87, p = .001$.

Table 5 presents the summary from separate univariate ANOVAs based on trimmed means (Yuen’s modified $t$-test; Wilcox, 2012) followed by the effect size. These results are provided by the R-package WRS2 (Version 1.1.4; Mair & Wilcox, 2020). The effect size of each latent variable contribution to the gender differences was estimated using Wilcox and Tian (2011) explanatory measure of effect size, which is a robust alternative to Cohen’s $d$ that allows heteroscedasticity. Similarly, to Pearson’s correlation, the explanatory measure of effect sizes 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female n: 86</th>
<th>Male n: 68</th>
<th>$M_{\text{diff}}$ [95% CI]</th>
<th>$M_{\text{diff}}$</th>
<th>$r$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSC</td>
<td>-0.37</td>
<td>0.25</td>
<td>-0.62 [-0.94, -0.29]</td>
<td>3.81</td>
<td>0.46</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICA</td>
<td>0.25</td>
<td>-0.48</td>
<td>0.73 [0.41, 1.05]</td>
<td>4.47</td>
<td>0.50</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GMS</td>
<td>-0.16</td>
<td>0.17</td>
<td>-0.34 [-0.63, -0.05]</td>
<td>2.32</td>
<td>0.28</td>
<td>0.023</td>
</tr>
<tr>
<td>ASA</td>
<td>0.24</td>
<td>-0.41</td>
<td>0.65 [0.34, 0.96]</td>
<td>4.10</td>
<td>0.48</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>pGr</td>
<td>2.77</td>
<td>2.36</td>
<td>0.41 [-0.19, 1.01]</td>
<td>1.37</td>
<td>0.17</td>
<td>0.176</td>
</tr>
</tbody>
</table>

Note: $M_{\text{diff}}$ = trimmed mean difference, $Y_t$ = test statistic, $r$ = explanatory measure of effect size, $pGr$ = prior math achievement.
MSC than female students, $M_{\text{diff}} = -0.62 [-0.94, -0.29], Y_t = 3.81, p < .001$, and the opposite is found for assignment anxiety, $M_{\text{diff}} = 0.65 [0.34, 0.96], Y_t = 4.10, p < .001$.

The relative contribution of each factor

To summarize, Table 5 demonstrates that ICA is a significant factor in explaining gender differences. However, in order to further investigate the factors that influence gender discrimination, a series of logistic mixed effect models are employed (as depicted in Table 6), to examine the disparities between gender and prior mathematics achievement, MSC, GMS, ICA, and ASA.

Nakagawa’s marginal and conditional $r^2$-squared values for mixed models, as previously mentioned, are estimated by the R-package performance (Version 0.10.4; Lüdecke et al., 2021). The marginal $r^2$-squared value is indicated in parentheses. Furthermore, the profile likelihood confidence intervals are provided by the R-package lme4 (Version 1.1.34; Bates et al., 2015).

First, gender is predicted by past mathematics achievement (Model 1). The analysis demonstrates that the estimations in the model are influenced by the mathematics course level, with marginal and conditional $r^2$-squared values of 0.03 and 0.12, respectively. However, the model fit is not statistically significant, $\chi^2(4) = 3.62, p = 0.46$, which indicates that gender differences cannot be explained by prior mathematics achievement alone.

In Model 2, the self-concept-like factors (MSC and GMS) are added as predictors, which according to the likelihood ratio test (see $\Delta$Deviance in Table 6) explained considerably more of the gender differences, $\chi^2(2) = 19.55, p < .001$. In the last model (Model 3), the mathematics anxiety factors (ICA and ASA) are added as predictors. This resulted in an increased in explanation power of gender differences according to the likelihood ratio test, $\chi^2(2) = 16.60, p < .001$. Moreover, Model 3 demonstrates that only in-class anxiety remained a significant discriminator even when controlled for the other variables.

Certain confidence intervals (CIs) almost comprise one, and thus should be interpreted with caution. For instance, when all other predictors are maintained constant, the odds of male students achieving mathematics grade A are 84% less likely than those with a mathematics grade C. Furthermore, the odds of male students achieving prior mathematics grade E are 3.89 times greater than those of the reference group with a grade C. Nonetheless, the odds ratio confidence intervals are exceedingly large for grade E as a predictor, indicating that it is not a reliable estimate.
However, one parameter estimates demonstrate strong statistical significance. Model 3 demonstrates that for each unit increase in ICA (in-class anxiety), the odds are 57% less likely for it to be a male student.

Discussion and conclusion

Only a few research studies on general mathematics self-efficacy and anxiety beliefs have dealt with extreme multicollinearity problems due to difficulties in operationalization. For instance, in some large-scale studies (e.g. Marsh et al., 2019), intercorrelations between factor-related items are frequently unreported, raising concerns about the potential impact of extreme multicollinearity on results. This article presents a study that addresses potential issues with extreme multicollinearity before conducting factor analysis.

The EFA results revealed two self-concept-like constructs, representing facets of students’ mathematics self-concept, and two facets of mathematics anxiety. One self-concept-like factor indicates students’ future-oriented perception of their mathematics competence, labeled generalized mathematics self-efficacy, while the other reflects their overall confidence in mathematics (“I am good at mathematics”). The mathematics anxiety factors capture students’ fear of providing incorrect answers in class and their anxiety and stress in evaluative situations, such as failing to complete assignments on time or struggling with homework.

Students’ mathematics self-concept beliefs strongly correlated with both mathematics assignment anxiety ($r = −0.60$) and generalized mathematics self-efficacy ($r = 0.61$), indicating a close association between confidence in mathematics and overall competence evaluation, consistent with previous research (e.g. Goetz et al., 2010; Holenstein et al., 2022; Marsh et al., 2019). Mixed models, with course level as a random effect, suggested that past mathematics achievement more strongly predicted the mathematics self-concept factor than generalized mathematics self-efficacy. However, unlike findings by Marsh et al. (2019), there was no indication that generalized mathematics self-efficacy and mathematics self-concept were empirically identical constructs.

Consistent with prior research (e.g. Goetz et al., 2013), this study found that both female and male students had comparable prior mathematics achievement. However, despite this similarity, female students reported lower mathematics self-concept and generalized mathematics self-efficacy, along with higher levels of in-class anxiety and assignment anxiety compared to male students. Notably, in-class anxiety emerged as a significant discriminator, exerting a substantial effect on explaining gender differences even after controlling for other variables in a logistic mixed model with course level as a random effect.

One explanation for the gender differences in in-class anxiety could be attributed to gender-stereotype effects (Goetz et al., 2013). However, this finding may also stem from students’ perceptions of success in mathematics, where the correct answer is seen as the ultimate proof of understanding, potentially leading to feelings of isolation among those facing difficulties. Therefore, differences in classroom settings, including the behavior of teachers, may impact students reported in-class anxiety, thus suggesting the need for further exploration of this phenomenon.

Despite its strengths, such as thorough data preparation and addressing multicollinearity, this study also has limitations. First, the sample size is relatively small, drawn from two studies, limiting generalizability. To enhance generalizability, multiple samples from diverse classroom settings are necessary. Second, regarding the reliability of the final factor model, each factor comprises only three items, and one factor includes an item with a significant cross-loading, which may impact the model’s reliability. However, as this study is exploratory, these limitations were not deemed problematic. The four factors explain 65% of the variance in the data and represent items with both high loadings and high communalities, which are indicators of a strong factor model (Costello & Osborne, 2005; Thompson, 2004).
Despite significant limitations in sample size, this study offers valuable insights for future research. For instance, this study illustrates the significance of addressing highly overlapping variables before factor analysis. It makes it easier to understand the factor structure and avoids having to remove factors in regression models, as shown in studies by Pajares and Miller (1994) and Marsh et al. (2004). This insight holds relevance to all types of quantitative studies, regardless of their size, when facing challenges in operationalization.

Essentially, this study demonstrates the importance of examining self-efficacy beliefs within specific research contexts, especially considering their predictive power and interaction with other self-belief constructs. This is because problems with high levels of multicollinearity in data may result from subtle distinctions between questionnaire items that respondents struggle to interpret differently, highlighting the significance of the research context. Although a high level of multicollinearity poses challenges for quantitative analyses, it can also appear in qualitative analyses. However, it could manifest as overlapping themes and challenges in interpreting results, instead of very high correlation.

This study addresses highly correlated variables by carefully selecting one for removal. This decision considers both quantitative criteria and qualitative evaluations of questionnaire items to identify redundant items within the specific research context. This methodological approach could be especially relevant in discussions about gender disparities, as the findings are closely tied to the specific context of the study (Else-Quest et al., 2010). Moreover, this approach provides opportunities for future exploration and a more profound comprehension of self-concept-like constructs.

Disclosure statement
The author reports there are no competing interests to declare.

ORCID
Erik Bergqvist http://orcid.org/0000-0002-7587-7152

Data availability statement
The anonymized version of data used in this study is available upon request.

References


