Journal of Adolescence 80 (2020) 220–232

Contents lists available at ScienceDirect

Journal of Adolescence

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

The longitudinal role of mathematics anxiety in mathematics development: Issues of gender differences and domain-specificity

Zhe Wang[a,*], Kaili Rimfeld[b], Nicholas Shakeshaft[b], Kerry Schofield[b], Margherita Malanchini[c,**]

[a] Department of Human Development and Family Studies, Texas Tech University, Lubbock, TX, USA
[b] Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[c] School of Biological and Chemical Sciences, Queen Mary University of London, London, UK

ARTICLE INFO

Keywords:
Mathematics anxiety
Mathematics development
Gender differences
Domain-specificity
Longitudinal
Adolescence

ABSTRACT

Introduction: Mathematics anxiety (MA) is an important risk factor hindering the development of confidence and capability in mathematics and participation in the science, technology, engineering, and mathematics workforce. The aim of the present study is to further our understanding of these relations in adolescence by adopting a threefold approach. First, we adopted a longitudinal design to clarify the temporal order in the developmental relations between (a) MA and mathematics achievement and (b) MA and mathematics self-perceived ability. Second, we investigated whether the developmental relations between MA and mathematics achievement/self-perceived ability differed between boys and girls. Finally, we explored the domain-specificity of MA by examining its role in foreign language (L2) learning.
Methods: Data were collected from 1043 Italian high school students. Students reported their anxiety, self-perceived ability, and school achievement in mathematics and L2 over two separate waves, one semester apart.
Results: Using multi-group cross-lagged panel analyses, we found that (a) mathematics achievement predicted MA longitudinally, whereas MA did not predict subsequent mathematics achievement; (b) there was a negative reciprocal relation between MA and mathematics self-perceived ability in male, but not female students; and (c) there were longitudinal relations between MA and L2 achievement and self-perceived ability above and beyond L2 anxiety.
Conclusions: These findings support the deficit view of the developmental relation between MA and mathematics achievement, highlight high school male students as a vulnerable group evincing vicious transactions between high anxiety and low self-efficacy in mathematics, and reveal the importance of internal cross-domain comparison processes in MA development.

1. Introduction

Mathematics anxiety (MA) refers to the experience of fear and apprehension in the anticipation or in the midst of mathematics-related activities (Hopko, Mahadevan, Bare, & Hunt, 2003). MA has been proposed as one important risk factor in the development of lower subjective confidence and objective capability in mathematics, as well as lower rates of participation in the science, technology,

* Corresponding author. Department of Human Development and Family Studies, 1301 Akron Ave Box, 41230, Lubbock, TX, USA.
** Corresponding author. School of Biological and Chemical Sciences, G.E. Fogg Building, Office 2.02, Mile End Road, London, E1 4NS, UK.
E-mail addresses: zhe.wang@ttu.edu (Z. Wang), m.malanchini@qmul.ac.uk (M. Malanchini).

https://doi.org/10.1016/j.adolescence.2020.03.003
Received 19 July 2019; Received in revised form 5 March 2020; Accepted 9 March 2020

engineering, and mathematics (STEM) workforce (Dowker, Sarkar, & Looi, 2016; Ramirez, Shaw, & Maloney, 2018). The important implications of these negative relations warrant further investigation into the development of MA and its dynamic relations with these mathematics-related outcomes. Grounded in the *Control-Value Theory* of achievement emotion (Pekrun, 2006), the current work addressed three under-investigated areas of research in adolescence, namely: (1) the directionality of the relation between MA and mathematics achievement and self-perceived ability, (2) gender differences in these developmental associations, and (3) the relation between MA development and the development of confidence and ability in other academic domains.

1.1. The developmental relations between MA, achievement and self-perceived ability

Theoretically, the *Control-Value Theory* proposes a dynamic reciprocal relation between achievement emotion and academic performances, and between achievement emotion and control/value beliefs over time (Pekrun, 1992, 2006). Achievement emotion refers to emotions that are tied directly to achievement activities or outcomes. For example, some students may experience MA as they sit in a mathematics class (i.e., activity emotion), others may experience MA primarily in anticipation of an upcoming mathematics exam (i.e., outcome emotion). According to the *Control-Value Theory* (Pekrun, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007), individuals' emotional experiences in a particular academic domain are shaped by their perceived control over and utility value of the learning activities and outcomes in that domain. In turn, these domain-specific achievement emotions affect students' learning behaviors and academic outcomes, which ultimately influence their perceived control over their behaviors and outcomes in that domain. In this way, achievement emotions, self-perceived control, and academic achievement transact reciprocally (Ahmed, Minnaert, Kuyper, & van der Werf, 2012; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). Pertinent to the current investigation, poor mathematics achievement and a perception of low mathematics ability imply less control over the likelihood of success in mathematics, which may cause anxiety about future outcomes. This is consistent with the *deficit* view of MA, which posits that students with low mathematics ability experience more obstacles in their mathematics education, which may translate into more negative emotions towards mathematics (Tobias, 1986). In turn, MA may lead to a sequence of actions such as avoidance of learning and practice in mathematics, which will result in poor achievement and reinforce a sense of lack of control. This aligns well with the *debilitating* view of MA, which posits that MA causes mathematics deficits through competing for cognitive resources (Ashcraft & Moore, 2009; Beilock & Carr, 2005; Suárez-Pellicioni, Núñez-Peña, & Colomé, 2014), and fostering behavioral avoidance in mathematics learning (Hembree, 1990).

Empirically, studies have shown that higher MA is associated with lower mathematics performance and lower perceived control in mathematics (Hembree, 1990; Ma, 1999; Namkung, Peng, & Lin, 2019), relations that are observed across different cultures (Kalaycioğlu, 2015; Lee, 2009). Yet, most existing studies examining these relations rely on correlational and cross-sectional designs, failing to clarify the direction of these observed relations. Among the few studies that utilized a longitudinal design, findings are mixed with respect to the directions of the relations.

Regarding the relation between MA and mathematics achievement, several studies in early elementary school students provide evidence supporting the debilitating view of MA and suggest that the prediction flows from MA to mathematics achievement (Cargnelutti, Tomasetto, & Passolunghi, 2017; Ching, 2017; Vukovic, Kieffer, Bailey, & Harari, 2013). Two of these longitudinal studies, one in a group of Chinese students and the other in a group of U.S. students, found that MA in Grade 2 predicted lower mathematics performance and lower gains in mathematics achievement in Grade 3 (Ching, 2017; Vukovic et al., 2013). However, these two studies did not examine the effect of early mathematics achievement on subsequent development of MA. Using a cross-lagged panel model that allowed the examination of effects in both directions, one study in a group of Italian students found that MA in Grade 2 predicted mathematics achievement in Grade 3, whereas mathematics achievement in Grade 2 did not predict MA in Grade 3, supporting the debilitating view of MA (Cargnelutti et al., 2017). However, a recent study that employed a similar design found that MA and mathematics achievement mutually predicted each other across six months in a group of first and second graders in the U.S. (Gunderson, Park, Maloney, Beilock, & Levine, 2018). Two other longitudinal studies examined the relation between MA and mathematics achievement in adolescence. In support of the deficit view of MA, one longitudinal study in a large nationally representative sample of middle and high school students in the U.S. reveals that the prediction flows from mathematics achievement to MA (Ma & Xu, 2004). Using a multi-wave longitudinal design, this study found that mathematics achievement consistently predicted subsequent development in MA, but MA did not predict subsequent achievement development. However, another study in a group of German secondary school students found evidence supporting the reciprocal view of academic emotion and achievement outcomes (Pekrun et al., 2017). Specifically, negative reciprocal relations were found between negative mathematics emotions, including MA, and mathematics achievement across a four-year span (Pekrun et al., 2017).

Regarding the relation between MA and perceived control in mathematics, the most commonly examined aspect of perceived control is mathematics self-perceived ability or self-efficacy (Hembree, 1990), which refers to an individual's belief in successfully executing actions required for the desired outcomes (Bandura, 1977). Hembree's meta-analysis reported strong negative correlations (~ -0.7) between MA and self-perceived ability in mathematics based on cross-sectional studies (Hembree, 1990). Similarly, more recent cross-sectional studies revealed modest to moderate negative associations between MA and mathematics self-perceived ability in children from non-western cultures across all school ages (Akin & Kurbanoglu, 2011; Justicia-Galiano, Martin-Puga, Linares, & Pelegrina, 2017; Lavasani, Hehazi, & Varzaneh, 2011; Nie, Lau, & Liau, 2011).

Two studies used longitudinal data to examine the relation between MA and mathematics self-perceived ability. In one study of middle school students, it was found that higher perception of one's own mathematics ability predicted lower subsequent MA (Meece, Wigfield, & Eccles, 1990). However, this study did not examine the predictive effect of MA on subsequent self-perceived ability. One other study in a group of middle school students found supporting evidence for the *Control-Value Theory* (Ahmed et al., 2012), showing that students with higher MA developed lower mathematics self-perceived ability, and students with higher self-perceived ability developed lower MA one year later.

To summarize, the limited literature presents mixed findings regarding the directions of the associations between MA and mathematics achievement/self-perceived ability. One possible reason for this inconsistency is that several longitudinal studies were not designed in a way in which effects in both directions can be tested (Meece et al., 1990; Vukovic et al., 2013). In the present study, by measuring all constructs at two different time points, we were able to examine effects in both directions. Additionally, different studies focused on different educational stages. It is possible that the developmental relation may change across educational stages, resulting in differences between studies on childhood (e.g., Gunderson et al., 2018; Krinzinger, Kaufmann, & Willmes, 2009) and studies on adolescent years (e.g., Ahmed et al., 2012; Ma & Xu, 2004). While most longitudinal studies were conducted in elementary and middle school students, only one study investigated the link between MA and mathematics achievement in high school students (Ma & Xu, 2004), and none has examined the link between MA and mathematics self-perceived ability in high school students. Thus, the present study first aimed to contribute to this under-investigated area by examining the longitudinal relations between MA and mathematics achievement/self-perceived ability in high school.

1.2. Gender specificity

The second aim of the present study was to explore possible gender differences in the developmental relations between MA and mathematics achievement/self-perceived ability. The *Control-Value Theory* argues that the relations between perceived control and achievement emotions should be structurally equivalent in male and female students (i.e., achievement emotions depend on subjective control and value to the same degree in both genders), although the mean levels of perceived control and achievement emotions may differ between the two gender groups (Pekrun, 2006). Most existing studies on gender differences focused on mean level differences. Results from meta-analyses suggest that at the mean level, male students are more confident in their mathematics ability and less anxious about mathematics compared to female students, but female students' mathematics performance are on par with their male peers (Else-Quest, Hyde, & Linn, 2010; Hyde, 2014). Very few studies have explored gender differences in the strengths of the relations between MA and mathematics achievement/self-perceived ability. Among these studies, one found that MA was related to arithmetic calculation more strongly in adult men than adult women (Miller & Bichsel, 2004). Similarly, in a group of middle to high school students, one study found that mathematics achievement predicted MA more consistently in male students than in female students (Ma & Xu, 2004). However, two other studies in groups of younger students showed the opposite pattern, where a negative correlation between MA and math performance was found in middle school girls but not boys (Devine, Fawcett, Szűcs, & Dowker, 2012; Geary et al., 2019). Given that only a few studies explored gender differences in the relations between MA and mathematics development and presented mixed findings, the present study aimed to further explore this under-investigated question.

1.3. Domain specificity

The third aim of the present study was to investigate whether the relations between MA and mathematics achievement/self-perceived ability are specific to the mathematics domain. In other words, is MA associated with achievement and motivation in other academic domains? The *Control-Value Theory* posits that achievement, perception of one's control (or ability), and achievement emotions are primarily linked with one another in a domain-specific manner (Pekrun, 2006). For example, MA is a mathematics domain-specific emotion that is theorized to be unrelated to achievement and motivation in other academic domains. Some empirical evidence supports this argument by showing a minimal correlation between MA and verbal achievement (Hembree, 1990).

While the *Control-Value Theory* focuses on the domain-specific relations, an alternative theory on achievement emotion, the G*eneralized Internal/External Frame of Reference Model* (Möller, Möller-Kalthoff, Helm, Nagy, & Marsh, 2016), focuses on the cross-domain relations. This theory posits that the development of achievement emotions in one domain is also subject to the influence of achievement and self-perceived ability in another domain, and that achievement emotion in one domain may also affect achievement and self-perceived ability development in another domain (Möller et al., 2016). According to the *Generalized Internal/External Frame of Reference Model*, individuals engage in both external (comparing one's achievement in one domain to others' achievement in the same domain) and internal (comparing one's achievement in one domain to one's achievement in another domain) comparison processes in the formation of emotion and motivation in learning (Arens, Becker, & Möller, 2017; Möller, Retelsdorf, Köller, & Marsh, 2011). If a student has better verbal achievement than mathematics achievement, or if a student is more confident with his/her verbal ability compared to his/her mathematics ability, an internal comparison between the two domains may lead to more anxiety about the lower-achieving or less confident domain (i.e., mathematics; Arens et al., 2017; Möller et al., 2011). It is also possible that higher MA may, in turn, steer students away from mathematics learning and toward language learning, which would result in further decreases in achievement and self-perceived ability in mathematics and increases in these constructs in the verbal domain, a proposition yet to be tested. One recent cross-sectional study incorporated within-domain and cross-domain relations into the same model, and found that test anxiety in mathematics was positively associated with verbal achievement and self-perceived ability (Arens et al., 2017). The G*eneralized Internal/External Frame of Reference Model* (Arens et al., 2017; Möller et al., 2011) argues that the positive correlations between test anxiety in mathematics and achievement and self-perceived ability in the verbal domain were driven by an internal comparison process.

Largely missing from the literature are studies that incorporate anxiety, self-perceived ability, and achievement from more than one domain in a longitudinal design, which allows for the investigation on how these cross-domain relations evolve over time. Therefore, the third aim of the present study was to address this major gap by investigating the relations between anxiety and achievement/self-perceived ability in the domains of mathematics and foreign language learning (L2) simultaneously in a longitudinal design. We focused on investigating whether L2 achievement/self-perceived ability predict subsequent development in MA, and whether MA predicts subsequent development in L2 achievement/self-perceived ability.

We chose to examine L2 as the other domain for two main reasons. First, in line with the G*eneralized Internal/External Frame of Reference Model* and the dimensional comparison model (Arens et al., 2017), a verbal domain, such as foreign language learning, stands in contrast to a quantitative domain. Several studies have shown negative correlations between L2 self-concept and mathematics achievement (Marsh, Kong, & Hau, 2001; Möller, Streblow, Pohlmann, & Kö;ller, 2006; Xu et al., 2013), suggesting that students may engage in internal comparison processes in these two domains. Second, there is a well-established literature on L2A. Valid and reliable measures of L2A have been developed (Csizér & Kormos, 2009; Horwitz, Horwitz, & Cope, 1986). There have also been numerous studies on understanding the underpinnings of L2A, such as the sources of variability in L2A and effects of L2A on L2 learning and achievement (for a review, see Horwitz, 2010). In general, students with higher L2A have lower L2 proficiency, are less willing to communicate in L2, and tend to underestimate their ability in L2 communication (MacIntyre & Charos, 1996; MacIntyre, Noels, & Clément, 1997; Sparks, Patton, Ganschow, & Humbach, 2009).

1.4. Aims and hypotheses

To summarize, the present study aimed to advance our understanding of the developmental relations between MA and mathematics achievement and self-perceived ability by 1) utilizing a longitudinal design to investigate the direction of effects; 2) examining gender as a moderator to explore whether and how MA relates to mathematics achievement/self-perceived ability differentially between male and female students; and 3) incorporating another academic domain (L2) to investigate the domain specificity of the MA – mathematics achievement/motivation relations. We formulated our hypotheses based on the *Control-Value Theory*. We hypothesized that there would be negative reciprocal relations between MA and mathematics achievement, as well as between MA and mathematics self-perceived ability (aim 1). We hypothesized that there would be no difference (i.e., structural equivalence) between male and female students in the relations between MA and mathematics achievement/self-perceived ability (aim 2). Finally, we hypothesized that MA and achievement/self-perceived ability in L2 would not be longitudinally related (aim 3). To exclude possible confounding of domain-general processes, general anxiety and general cognitive abilities, including verbal ability, non-verbal reasoning ability, and working memory, were controlled for in the investigation of the three main research questions.

In the present study, we focused on investigating the above three questions in high school students. High school is a stage when important educational and occupational choices are made. As students decide which higher education or career trajectory to embark on, they may need to consider and critically evaluate their ability level in a domain, their self-confidence in future success in that domain, as well as their potential emotional well-being pursuing careers in that domain (Ganley, George, Cimpian, & Makowski, 2018; Hackett, 1985; Moakler & Kim, 2014), which renders these constructs and their relations salient in this developmental stage. Additionally, high school is a stage when students' academic motivation becomes more differentiated among different domains (Bong, 2001). As students pitted their emotions, motivations, and ability levels in one domain against another to decide on an education or career path, the internal cross-domain comparison process may also be particularly salient during high school. Yet, few studies have investigated issues such as the directions of the effect, gender-specificity, and domain-specificity in this important educational stage in a longitudinal design. Therefore, by focusing on high school students, the current study would contribute to our growing understanding of how the relations between achievement emotion and achievement/self-perceived ability within and across domains develop in this critical period of time.

2. Methods

2.1. Participants

Participants in the present analyses were part of the Multi-Cohort Investigation into Learning and Educational Success (MILES) study (www.projectmiles.com). Students from three opportunistically selected state high schools (one Licee, more academically oriented, and two Technical schools) in the Province of Milan were invited to participate in the study in February (wave1) and June 2016 (wave2). These two data collections took place at the end of the autumn and spring semesters, respectively. In total, 1083 students participated in wave1, of whom 444 students also participated in wave2. The present study analyzed data from the 1043 (53% female) students who provided data on at least one wave1 measure. Multivariate analysis of variance suggested that there were no significant differences between students who participated only in wave1 and those who participated in both waves in any of the main study variables (i.e., MA, L2A, mathematics achievement, L2 achievement, mathematics self-perceived ability, and L2 self-perceived ability in both waves; Wilk's $\lambda = 0.98$, $p = 0.973$). The age of the students ranged from 14 to 21 years in wave1 ($M = 15.86$, $SD = 1.52$). The distribution of the sample across schools and grades are shown in supplementary materials Table S1. MILES received approval from the teachers' and parents' committee of every school, as well as ethical approval from Goldsmiths University of London.

2.2. Measures

All measures were administered online in Italian. Prior to the first wave of data collection, all measures were translated from English to Italian and piloted on a sample of 70 students in Milan. The distribution and internal validity of each construct as well as the correlations between constructs were comparable to those obtained with the validated English versions of the measures administered to English-speaking samples. The translated measures are included in the supplementary materials (Table S2 – S5).

2.2.1. Mathematics anxiety

The Abbreviated Mathematics Anxiety Scale (Hopko et al., 2003) was used to measure MA. Students were asked to rate on a 5-point scale (1 = not at all; 5 = very much) how anxious they feel in nine mathematics-related situations, such as "taking an examination in a math course". This scale showed good internal reliability, with Cronbach's α being 0.85 in wave1 and 0.88 in wave2. A MA score was calculated by taking the mean of all item scores. Higher scores represented higher MA.

2.2.2. Foreign language (L2) anxiety

The seven anxiety items from the L2 motivation scale developed by Csizér and Kormos (2009) was used to measure L2 anxiety. Students were asked to rate on a 5-point scale (1 = not true at all; 5 = very true) to what extent each statement applied to them. Students who were studying more than one foreign language were asked to think about their main foreign language when answering the questions. An example item was "I would feel uneasy speaking the foreign language I study with/to a person who speaks that language". L2 anxiety score was calculated by taking the mean of all item scores, with higher scores indicating higher L2 anxiety. Cronbach's α for the L2 anxiety scale was 0.88 in wave1 and 0.90 in wave2.

2.2.3. Mathematics self-perceived ability

To measure self-perceived ability in mathematics, students were asked to rate how good they believed they were at three mathematics-related activities on a 5-point scale (1 = not good at all; 5 = very good; Spinath, Spinath, Harlaar, & Plomin, 2006). The three activities were solving number and money problems, doing mathematics in one's head, and multiplying and dividing numbers. Cronbach's α for this scale was 0.77 and 0.80 respectively in wave1 and wave2. The three items were averaged to obtain the mathematics self-perceived ability score. Higher scores represented higher self-perceived ability in mathematics.

2.2.4. Foreign language (L2) self-perceived ability

L2 self-perceived ability was measured by asking students how good they believed they were at L2 reading, writing, and grammar on a 5-point scale (1 = not good at all; 5 = very good; Spinath et al., 2006). Note that the original scale administered to English speakers measures self-perceived ability in reading, writing, and spelling. We adapted spelling to grammar due to inherent differences among languages. Cronbach's α was 0.81 in wave1 and 0.83 in wave2. L2 self-perceived ability score was obtained by averaging the three item scores, with higher scores representing higher self-perceived ability in L2.

2.2.5. Mathematics and foreign language achievement

Students self-reported their school grades in mathematics and L2 classes in both waves. Scores ranged from 4 to 10. Higher scores indicated better achievement performance.

2.3. Covariates at Wave1

2.3.1. General anxiety

Students' general anxiety was measured using the Generalized Anxiety Disorder 7-item scale (Löwe et al., 2008). Students were asked how often they have been bothered by a number of problems during the past two weeks using a 4-point scale (1 = not at all; 4 = nearly every day). An example item was "feeling nervous, anxious, or on edge". Cronbach's α was 0.89 for this scale.

2.3.2. Raven's Progressive Matrices

Students' non-verbal reasoning was measured using the short version of the Raven's Progressive Matrices test (Raven, Court, & Raven, 1996). This test consists of 30 incomplete matrices. Students were asked to identify the missing part of each matrix by choosing from eight possible options. The total number of correct responses was used to indicate each student's performance on this task. This test is shown to have good internal reliability (Haworth et al., 2007). Detailed information about the online administration of this test can be found at http://www.teds.ac.uk/datadictionary/studies/webtests/16yr_raven_test.htm.

2.3.3. Corsi Block

An online version of the Corsi Block-Tapping Task (Pagulayan, Busch, Medina, Bartok, & Krikorian, 2006) was used to measure students' spatial working memory. Nine cubes are shown on the screen. The cubes light up, turning yellow for 1 s in a patterned sequence, with a 1-s interval between each cube. Participants were asked to reproduce the sequence by clicking on the cubes using the mouse. The test increases in difficulty level, ranging from a four-cube sequence to a nine-cube sequence. Two sequences were administered within each difficulty level, and the test was terminated when both sequences in the same level were reproduced incorrectly. The total number of correctly reproduced sequences was the score used in the following analyses. This test is shown to have good internal reliability (Busch, Farrell, Lisdahl-Medina, & Krikorian, 2005).

2.3.4. Italian Vocabulary test

To measure students' native language vocabulary, the MILES team developed the first self-administered online vocabulary test in Italian. Participants were presented with one Italian word at a time, and were asked to choose a synonym of that word out of six

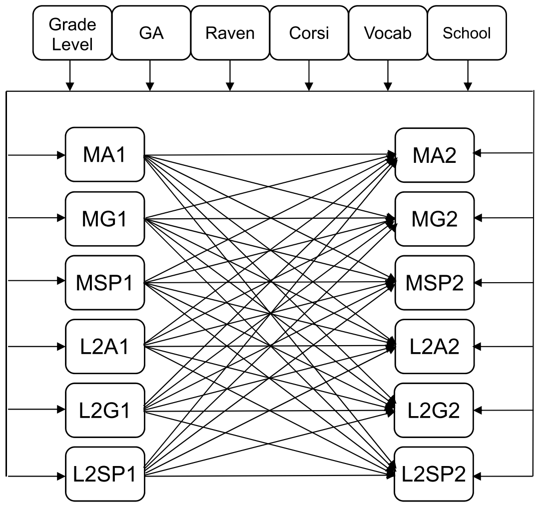

Fig. 1. Cross-lagged panel model. MA = mathematics anxiety, MG = mathematics achievement, MSP = mathematics self-perceived ability, L2A = foreign language anxiety, L2G = foreign language achievement, L2SP = foreign language self-perceived ability; GA = general anxiety; Raven = Raven's Progressive Matrices; Corsi = Corsi Block; Vocab = Italian Vocabulary. 1 and 2 in the variable labels indicate wave1 and wave2, respectively. Concurrent correlations and residuals are not shown for simplicity.

possible options. Eighty words were presented. Words with higher frequency (i.e., more commonly occurring in written Italian) were presented before words with lower frequency. The total number of correct answers was used to measure each student's performance. The test showed good test-retest reliability ($r = 0.60$) over a four month period, and good criterion validity, correlating at $r = 0.46$ with the Raven's Progressive Matrices test, at $r = 0.20$ with the Corsi Block test, and at $r = 0.30$ with students' L2 achievement.

2.4. Analytic strategies

SPSS 24 (IBM Corp, 2016) was used to obtain the descriptive statistics and correlations between measures. To investigate the main research questions, multi-group cross-lagged panel models were run in Mplus 8 (Muthén & Muthén, 1998–2017). Using a cross-lagged panel model (see Fig. 1), we were able to examine the strengths of the within-domain cross-lagged relations (e.g., MA in wave1 predicts mathematics achievement in wave2) as well as the cross-domain cross-lagged relations (e.g., MA in wave1 predicts L2 achievement in wave2), independent of the stability of each construct and their concurrent correlations.

To examine gender differences in these longitudinal relations, we conducted multi-group analyses in which a certain path is constrained to be equal across gender groups. If a model with such a constraint fits significantly worse than the baseline model with no such constraint, it indicates a difference in the strength of the relation between the gender groups. Specifically, we first ran a full model (Fig. 1) in which all paths were free to be different across gender groups. We then ran thirty-six nested models to examine whether each of the thirty-six predictive paths (i.e., six stability paths, twelve within-domain cross-lagged paths, and eighteen cross-domain cross-lagged paths) differ between the two gender groups. Finally, we ran a model in which all predictive paths suggested to be equal in the above steps were constrained to be equal across groups. Parameter estimates from this final model are presented.

The chi-square difference test was used to compare models. A significant chi-square difference test indicated that the nested model fit worse than the full model. Due to the large number of between-group comparisons, statistical significance was defined to be a p value smaller than the pre-specified type I error rate of 0.05 after Holm-Bonferroni correction (Holm, 1979). The fit of each model was also evaluated based on the Comparative Fit Index (CFI; Bentler, 1990) and the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990). A CFI above 0.90 and a RMSEA under 0.08 indicate acceptable model fit. Concurrent correlations and predictive effects of covariates (including school, grade level, general anxiety, and performances on Raven's Progressive Matrices, Corsi Block, and Italian Vocabulary tests) were included in all models but are not presented in the main results section for simplicity. These estimates are presented in the supplementary materials.

Before running cross-lagged panel models, we tested the assumptions of normality, homoscedasticity, and missing completely at random (MCAR) using the procedure suggested by Jamshidian, Jalal, and Jansen (2014). Results showed that the assumptions of homoscedasticity and MCAR held ($p = 0.092$) while the assumption of normality was violated ($p < 0.000$). Therefore, bias-corrected bootstrap standard errors and confidence intervals were estimated, and missing values were handled using full information maximum likelihood in all models.

3. Results

Descriptive statistics for the main study variables are shown in Table 1. Scores for each variable were distributed widely across their respective scales. Descriptive statistics are also presented separately for boys and girls in supplementary materials Table S6.

Table 1
Descriptive statistics of main study variables.

	N	Mean	Standard Deviation	Skewness	Kurtosis	Median	Minimum	Maximum
Mathematics anxiety1	1043	2.47	0.74	0.38	−0.07	2.44	1.00	5.00
Mathematics anxiety2	430	2.40	0.81	0.51	0.23	2.33	1.00	5.00
Mathematics achievement1	1035	6.50	1.41	−0.00	−0.61	7.00	4.00	10.00
Mathematics achievement2	433	6.68	1.30	0.00	−0.35	7.00	4.00	10.00
Mathematics self-perceived ability1	1043	3.51	0.82	−0.66	0.46	3.67	1.00	5.00
Mathematics self-perceived ability2	429	3.45	0.84	−0.61	0.72	3.33	1.00	5.00
L2 anxiety1	1036	2.71	1.02	0.26	−0.72	2.67	1.00	5.00
L2 anxiety2	427	2.71	1.09	0.28	−0.76	2.67	1.00	5.00
L2 achievement1	1035	6.70	1.23	0.04	−0.40	7.00	4.00	10.00
L2 achievement2	433	6.90	1.19	0.08	−0.26	7.00	4.00	10.00
L2 self-perceived ability1	1036	3.40	0.82	−0.49	0.37	3.33	1.00	5.00
L2 self-perceived ability2	427	3.37	0.92	−0.59	0.15	3.33	1.00	5.00

Note. N = effective sample size for each variable. L2 = Foreign language. 1 and 2 in the variable labels indicate wave1 and wave2, respectively.

Table 2 presents correlations between the main study variables. Compared to male students, females had higher MA, L2A, L2 self-perceived ability, and L2 achievement, lower mathematics self-perceived ability, and comparable mathematics achievement. Within each domain, anxiety was negatively associated with achievement and self-perceived ability, and self-perceived ability was positively associated with achievement. For cross-domain correlations, achievement in the two domains were moderately positively correlated, whereas anxiety and self-perceived abilities in the two domains were modestly correlated.

To examine the main research questions, we compared model fit indices. The fit indices for all models are shown in supplementary materials Table S7. Results from model comparison suggested that, with the exception of the reciprocal paths between MA and mathematics self-perceived ability, all other predictive paths can be fixed to be equal across gender groups without worsening the model fit. As such, all predictive paths other than the reciprocal paths between MA and mathematics self-perceived ability were fixed to be equal across gender groups in the final model. The final model has an adequate fit, χ^2 (*df*) = 44.31(34), RMSEA = 0.026, and CFI = 0.997. Path estimates and their 95% confidence intervals from the final model are shown in Table 3. Estimates for predictive effects of covariates and concurrent correlations between the main study variables are shown in supplementary materials Tables S8 and S9.

In general, students' performances on the Raven's Progressive Matrices, Corsi Block, and Italian vocabulary test predicted the main study variables in wave1 but not wave2 (Table S8). Performance on the Raven's Progressive Matrices predicted constructs in the mathematics domain not in the L2 domain, whereas performance on the Italian vocabulary test predicted constructs in the L2 domain not in the mathematics domain. Corsi Block performance predicted constructs in both the mathematics and L2 domain.

Below, we focused on describing predictive paths directly pertaining to the three main research questions.

3.1. Aim 1: Directionality

With respect to the direction of the relation between MA and mathematics achievement, anxiety in wave1 did not predict achievement in wave2 (−0.08, *95% CI* = −0.26 to 0.08), but achievement in wave1 negatively predicted MA in wave2 (−0.09, *95% CI* = −0.14 to −0.03). With respect to the relation between MA and mathematics self-perceived ability, we found a reciprocal relation between anxiety and self-perceived ability in male students, not in female students. These findings are described in the section below.

3.2. Aim 2: Gender differences

With respect to gender differences, the differences in the longitudinal relations between MA and mathematics self-perceived ability between male and female students were the only paths moderated by students' gender. Male students who had higher MA reported lower self-perceived ability in mathematics in the subsequent wave (−0.29, *95% CI* = −0.55 to −0.08), and those who had lower self-perceived ability in mathematics reported higher MA a semester later (−0.20, *95% CI* = −0.36 to −0.03). In female students, MA and mathematics self-perceived ability did not predict one another longitudinally (wave1 MA to wave2 mathematics self-perceived ability: 0.07, *95% CI* = −0.12 to 0.08; wave1 mathematics self-perceived ability to wave2 MA: 0.06, *95% CI* = −0.07 to 0.19).

3.3. Aim 3: Domain specificity

With respect to cross-domain longitudinal relations, MA in wave1 did not predict L2 achievement in wave2 (0.08, *95% CI* = −0.07 to 0.22), but L2 achievement in wave1 significantly predicted MA a semester later (0.08, 95% CI = 0.01 to 0.14). In addition, there was a positive reciprocal relation between MA and L2 self-perceived ability (wave1 MA to wave2 L2 self-perceived ability: 0.18, *95% CI* = 0.02 to 0.33; wave1 L2 self-perceived ability to wave2 MA: 0.11, *95% CI* = 0.00 to 0.21).

Table 2
Correlations between main study variables.

	1. Sex	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2. Grade	0.02																
3. GA	−0.28*	0.15*															
4. Raven	0.12*	0.19*	−0.03														
5. Corsi	0.09	0.09	−0.04	0.28*													
6. Vocab	−0.02	0.26*	−0.08	0.42*	0.10												
7. MA1	−0.32*	−0.06	0.39*	−0.24*	−0.16*	−0.04											
8. MA2	−0.23*	−0.09	0.36*	−0.20*	−0.18*	−0.09	0.66*										
9. MG1	−0.05	−0.02	−0.19*	0.24*	0.08	0.24*	−0.28*	−0.29*									
10. MG2	0.03	0.03	−0.20*	0.26*	0.05	0.21*	−0.29*	−0.33*	0.72*								
11. MSP1	0.30*	−0.07	−0.30*	0.21*	0.17*	0.12*	−0.48*	−0.35*	0.25*	0.30*							
12. MSP2	0.28*	−0.11*	−0.29*	0.18*	0.19*	0.07	−0.42*	−0.38*	0.25*	0.32*	0.67*						
13. L2A1	−0.22*	0.09	0.29*	−0.02	0.00	−0.17*	0.18*	0.17*	0.00	−0.06	−0.08	−0.15*					
14. L2A2	−0.20*	0.11*	0.29*	−0.05	−0.01	−0.21*	0.10*	0.16*	−0.06	−0.06	−0.10	−0.15*	0.71*				
15. L2G1	−0.11*	−0.10*	−0.09	0.15*	0.01	0.30*	0.03	0.03	0.46*	0.38*	0.07	0.11*	−0.37*	−0.37*			
16. L2G2	−0.09	−0.05	−0.09	0.16*	0.01	0.31*	0.01	−0.01	0.40*	0.45*	0.13*	0.13*	−0.36*	−0.39*	0.77*		
17. L2SP1	−0.10*	−0.11*	−0.05	−0.03	−0.01	0.15*	0.08	0.10	0.09	0.12*	0.12*	0.14*	−0.51*	−0.45*	0.55*	0.58*	
18. L2SP2	−0.06	−0.10	−0.12*	0.05	−0.01	0.19*	0.05	0.00	0.15*	0.17*	0.15*	0.17*	−0.42*	−0.51*	0.47*	0.54*	0.69*

Note. MA = mathematics anxiety, MG = mathematics achievement, MSP = mathematics self-perceived ability, L2A = foreign language anxiety, L2G = foreign language achievement, L2SP = foreign language self-perceived ability, GA = general anxiety, Raven = Raven's Progressive Matrices, Corsi = Corsi Block, Vocab = Italian Vocabulary test. 1 and 2 in the variable labels indicate wave1 and wave2, respectively. * indicates statistical significance under the pre-specified type I error rate of 0.05.

Table 3
Final models: Path estimates and 95% confidence intervals.

Path	Coefficient	95% Confidence Intervals	Path	Coefficient	95% Confidence Intervals
MA1 → MA2	0.57*	(0.40, 0.71)	L2A1 → L2A2	0.65*	(0.52, 0.76)
MA1 → MG2	−0.08	(-0.26, 0.08)	L2A1 → L2G2	−0.03	(-0.13, 0.08)
MA1 → MSP2	(f): 0.07 (m): 0.29*	(-0.12, 0.27) (-0.55, −0.08)	L2A1 → L2SP2	−0.08	(-0.17, 0.01)
MA1 → L2A2	−0.07	(-0.28, 0.14)	L2A1 → MA2	0.12*	(0.04, 0.20)
MA1 → L2G2	0.08	(-0.07, 0.22)	L2A1 → MG2	0.02	(-0.11, 0.13)
MA1 → L2SP2	0.18*	(0.02, 0.33)	L2A1 → MSP2	−0.03	(-0.11, 0.05)
MG1 → MG2	0.57*	(0.50, 0.65)	L2G1 → L2G2	0.52*	(0.44, 0.63)
MG1 → MA2	−0.09*	(-0.14, −0.03)	L2G1 → L2A2	−0.08	(-0.16, 0.01)
MG1 → MSP2	0.02	(-0.04, 0.07)	L2G1 → L2SP2	0.05	(-0.04, 0.13)
MG1 → L2G2	0.10*	(0.02, 0.16)	L2G1 → MG2	0.08	(-0.03, 0.17)
MG1 → L2A2	0.02	(-0.07, 0.09)	L2G1 → MA2	0.08*	(0.01, 0.14)
MG1 → L2SP2	0.04	(-0.02, 0.11)	L2G1 → MSP2	0.01	(-0.06, 0.07)
MSP1 → MSP2	0.60*	(0.49, 0.71)	L2SP1 → L2SP2	0.66*	(0.51, 0.78)
MSP1 → MA2	(f): 0.06 (m): 0.20*	(-0.07, 0.19) (-0.36, −0.03)	L2SP1 → L2A2	−0.10	(-0.25, 0.08)
MSP1 → MG2	0.14*	(0.02, 0.29)	L2SP1 → L2G2	0.35*	(0.21, 0.48)
MSP1 → L2SP2	0.07	(-0.05, 0.19)	L2SP1 → MSP2	0.05	(-0.06, 0.16)
MSP1 → L2A2	−0.01	(-0.16, 0.14)	L2SP1 → MA2	0.11*	(0.00, 0.21)
MSP1 → L2G2	0.05	(-0.06, 0.18)	L2SP1 → MG2	0.04	(-0.11, 0.20)

Note. MA = mathematics anxiety, MG = mathematics achievement, MSP = mathematics self-perceived ability, L2A = foreign language anxiety, L2G = foreign language achievement, L2SP = foreign language self-perceived ability. 1 and 2 in the variable labels indicate wave1 and wave2, respectively. Covariates (school, grade level, general anxiety, and performances on Raven's Progressive Matrices, Corsi Block, and Italian Vocabulary tests), concurrent correlations, and residuals were estimated but are not shown in this table for simplicity. * indicates statistical significance under the pre-specified type I error rate of 0.05.

4. Discussion

Grounded in the *Control-Value Theory* of achievement emotion (Pekrun, 2006), the present study examined the longitudinal relations among anxiety, achievement, and self-perceived ability in mathematics and L2 in a sample of high school students, controlling for domain-general processes including general anxiety and several general cognitive abilities. Below, we discuss our findings related to each of the three primary aims outlined in the introduction.

The first aim of the present study was to examine the directionality of the relation between MA and mathematics achievement and self-perceived ability. According to the *Control-Value Theory*, we hypothesized negative reciprocal relations between these constructs. Regarding the relation between MA and mathematics achievement, findings did not fully support our hypothesis. We found significant prediction in only one direction: from mathematics achievement to subsequent MA. This finding suggests that the negative association between MA and mathematics achievement during high school is possibly attributable to the deficit process more than the debilitating process. Individuals with lower mathematics ability may experience more obstacles in mathematics learning, which make them more prone to MA development (Ma & Xu, 2004). Our finding is at odds with several studies in elementary school students that have shown predictive effect of MA on subsequent mathematics achievement (Cargnelutti et al., 2017; Gunderson et al., 2018), but is consistent with the one study in the literature that examined the longitudinal relation between MA and mathematics achievement in high school students (Ma & Xu, 2004). The divergence between studies on childhood and those on adolescence may reveal one possible reason why our finding did not fully support the *Control-Value Theory*: there may be a shift in the primary mechanism underlying the developmental relation between MA and mathematics achievement across educational stages. Both the deficit and debilitating processes appear to explain the association between MA and mathematics achievement (i.e., a bidirectional relation) in early educational stages. However, our finding, together with Ma and Xu (2004), suggests that the debilitating mechanism may play a decreasingly important role as individual differences in school achievement become increasingly stable over time (Rimfeld et al., 2018). That is, achievement emotions may be less predictive of subsequent school achievement in later educational stages after accounting for the stability of achievement. If so, even though effects flow in both directions when we consider the entire educational span, the dominant mechanism linking MA and mathematics achievement may vary in different educational stages. This postulation needs to be further investigated in multi-wave long-term longitudinal studies.

With respect to the longitudinal relation between MA and mathematics self-perceived ability, findings did not fully support our first hypothesis. While a reciprocal relation was found between MA and mathematics self-perceived ability in male students, such a relation was not observed in female students. This finding did not fully support the directionality proposed by the *Control-Value Theory*.

Related to the second aim of the present study, which was to explore gender differences in the MA-mathematics achievement/self-perceived ability relations, the differences between male and female students in the longitudinal relations between MA and mathematics self-perceived ability were the only gender difference revealed. This finding did not support the gender structural equivalence corollary of *Control-Value Theory* (Pekrun, 2006). This finding suggests that MA may have more negative implications in the development of self-perceived ability in male compared to female high school students. Despite gender similarities in mathematics achievement in all developmental stages (Else-Quest et al., 2010; Kersey, Braham, Csumitta, Libertus, & Cantlon, 2018), mathematics is stereotyped as a male domain in Western cultures (Ganley et al., 2018; Nosek, Banaji, & Greenwald, 2002). Many sociocultural factors, such as teachers' and parents' differential treatment of boys and girls in mathematics learning, may explain the development of unbalanced gender representation in mathematics (Jacobs, 1991; Jones & Wheatley, 1990). For boys who are less confident about their mathematics ability, it may be hard to reconcile the conflict between the low self-perceived ability in mathematics and the societal expectation that "mathematics is a boys' game". This belief of not being able to live up to expectations may further induce feelings of anxiety in boys. In addition, boys are shown to identify themselves with mathematics and feel obligated to perform well on mathematics more than girls (Cvencek, Meltzoff, & Greenwald, 2011; Nosek et al., 2002; Skaalvik & Skaalvik, 2004). Thus, a feeling of anxiety may be more threatening to a male student's self-identity and thus has more negative impact on their self-confidence in mathematics. To the contrary, female students are less likely to identify themselves with mathematics overall (Cvencek et al., 2011; Nosek et al., 2002), which may explain why MA did not predict self-perceived ability in mathematics in female students. This explanation is consistent with the identity threat literature which argues that individuals identifying themselves with a particular domain are more likely to be threatened by stress and anxiety experienced in that domain (Ramirez, McDonough, & Jin, 2017).

Note that our findings of the male-specific reciprocal relation between MA and mathematics self-perceived ability did not support the gender structural equivalence corollary of the *Control Value Theory*. This finding of gender structural nonequivalence may be unique to this developmental stage. As early as the first few years in elementary school, young girls may begin to acquire the stereotypical conception that females are inferior to males in mathematics, which further influences the emerging mathematics self-concept in young girls (Cvencek et al., 2011). Previous research has shown that gender differences in self-perception of mathematics competence emerge early (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Lindberg, Linkersdö;rfer, Ehm, Hasselhorn, & Lonnemann, 2013) and continue to widen across middle and early high school years (Fennema & Sherman, 1977; Skaalvik & Skaalvik, 2004). By late adolescent years, our finding together with previous meta-analytic studies (Else-Quest et al., 2010; Huang, 2013; Hyde, 2014) suggest that female students were consistently more anxious and less confident about their mathematics abilities than male students, in spite of similar mathematics achievement between the two gender groups. As such, it is possible that we did not observe the developing negative relation between MA and mathematics self-perceived ability in females because such a relation may develop much earlier for girls (Geary et al., 2019) and may be well-established and consequently leveled-off by late adolescence.

Finally, we examined the degree to which MA was related to achievement and self-perceived ability in another domain, namely L2. According to the domain-specificity corollary in the *Control-Value Theory*, we hypothesized that MA would not be associated with achievement/self-perceived ability in L2 longitudinally. However, results showed that MA positively predicted subsequent self-perceived ability in L2, and both achievement and self-perceived ability in L2 positively predicted subsequent MA, which did not support the *Control-Value Theory*. These findings largely corroborate the *Generalized Internal/External Frame of Reference Model* (Arens et al., 2017), according to which the positive cross-domain relations between MA and L2 achievement and self-perceived ability likely reflect the internal comparison process. Students who are confident about their L2 ability may worry more about their mathematics performance compared to their L2 performance. In turn, students who are anxious about their mathematics abilities may avoid mathematics and invest more time in other academic domains, such as L2, and become more confident in these other domains as a consequence. Compared to early educational stages, this internal comparison process may be particularly salient during high school when students' academic motivation becomes more differentiated among different domains, coinciding with the time when students start considering their higher education and career paths (Bong, 2001).

It is worth noting that the reverse relations between L2A and mathematics achievement/self-perceived ability were not observed. This finding is consistent with several previous studies that failed to observe any relation between emotion/motivation in the verbal domains and achievement in mathematics (Arens et al., 2017; Möller, Pohlmann, Köller, & Marsh, 2009; Schurtz, Pfost, Nagengast, & Artelt, 2014). This may suggest that compared to emotions/motivations in mathematics, emotions/motivations in the verbal domains are less influenced by the internal comparison processes, as students' experiences with verbal domains go beyond the school context, unlike their mathematics experiences which are mostly constrained within formal educational settings (Arens et al., 2017).

The present study has a number of limitations. First, data were available at two time points that were one semester apart. We were not able to explore long-term effects or potential transitions in the relations among anxiety, achievement, and self-perceived abilities across different developmental stages. Even among high school students, those who just entered high school (e.g., Grades 1 and 2 in high school) may differ from those who were near graduation (e.g., Grades 4 and 5). Second, the present study relied on students' self-reported school achievement. It is yet to be tested whether the present results are replicable using standardized mathematics achievement tests. Nonetheless, students' self-reported achievement scores were highly stable over time (above 0.7), indicating good reliability. In addition, the strengths of the correlations between MA and mathematics achievement are in line with findings from meta-analyses based on other standardized measures of mathematics achievement (e.g., Hembree, 1990), indicating that students' self-reported achievement has good criterion validity. A recent study in a large UK-representative sample of students found that self-reported achievement scores were nearly perfectly correlated with actual exam grades ($r > 0.95$ for mathematics, English, and Science; Rimfeld et al., 2018), supporting the validity of student self-reported achievement measures.

The present study has important implications for future research and education practices in the areas of MA. Theoretically, our findings did not fully support several corollaries of the *Control-Value Theory* of achievement emotion (Pekrun, 2006). First, we did not find a negative reciprocal relation between MA and mathematics achievement, or between MA and mathematics self-perceived ability in all students, failing to support the reciprocal linkages corollary of the *Control-Value Theory*. Second, while the relation between MA and mathematics achievement held equivalent between genders, the reciprocal relation between MA and mathematics self-perceived ability was much stronger in male students than in female students, a pattern that does not fully support the gender structural equivalence corollary. Finally, we found that the development of MA was subject to the influence of achievement/self-perceived ability not only in the mathematics domain, but also in a contrasting domain (i.e., L2). This finding corroborates the *Generalized Internal/External Frame of Reference Model*, and is at odds with the domain-specificity corollary of the *Control-Value Theory*. Overall, our findings share more similarities with prior work in samples of high school students (e.g., Ma & Xu, 2004) than with studies on early developmental stages (Cargnelutti et al., 2017; Geary et al., 2019; Gunderson et al., 2018). Together, these findings on the shifting directions of effect and varying degrees of gender- and domain-specificity highlight the possibility that the primary mechanism underlying the longitudinal relations between MA and mathematics achievement/self-perceived ability may vary across different educational stages. Compared to elementary and middle school students, high school students exhibit increasingly stable achievement (Rimfeld et al., 2018), more established gender beliefs (Gunderson, Ramirez, Levine, & Beilock, 2012), and more differentiated emotions and motivations among different academic disciplines (Bong, 2001). These crucial differences may account for the diverging findings in MA development among students in varying educational stages. The *Control-Value Theory* assumes that the developmental etiologies of achievement emotions are universal in all students, which does not take into account attributes unique to each developmental stage. Therefore, studies that utilize multi-wave long-term longitudinal designs that span across childhood, adolescence, and early adulthood are sorely needed to thoroughly examine these critical developmental issues and to refine existing theories on achievement emotions accordingly.

Practically, our findings of the longitudinal relation between MA and mathematics achievement suggests that treatment programs aiming to break this vicious cycle by addressing only the emotional component are unlikely to achieve long-lasting effects in later educational stages. Efforts may be better spent in later educational stages on addressing the mathematics deficit component in order to reduce levels of MA. Second, due to the disparity in STEM participation between men and women, much extant literature has focused on the negative consequences of gender stereotypes about mathematics on female learners (e.g., Beilock, Gunderson, Ramirez, & Levine, 2010; Smith, 2013). Our finding on gender differences in the MA – self-perceived ability relation suggests that such cultural expectations might be burdensome and constraining particularly for adolescent boys who do not perceive themselves as meeting such expectations. Finally, the finding on the cross-domain relations between MA and L2 achievement/self-perceived ability suggests that some students who are very good at mathematics may also develop MA if they perceive themselves to excel more in other academic domains compared to mathematics. In the long run, these students may avoid mathematics-related careers despite their capability of doing mathematics. This finding highlights the importance for treatment programs to take into consideration how unbalanced emotions, motivations, and abilities between domains might jeopardize the development of the lower-achieving domain.

Declaration of competing interest

The authors have no conflict of interest to declare.

Acknowledgements

We gratefully acknowledge all the students and teachers of the Istituto Tecnico S. Cannizzaro, Istituto Tecnico E. Matteri, Liceo C. Rebora and Liceo Scientifico E. Majorana in the city of Rho (Milan). We are particularly grateful to the principals and teachers Elena Barducci, Rita Loffredo, Nadia Boldrin, and Cosimo Morrone for their continuous support with the MILES project. During the preparation of this work, MM's work is partly supported by the 2017 American Psychological Foundation's (APF) David Wechsler Early Career Grant for Innovative Work in Cognition. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.adolescence.2020.03.003.

References

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences, 22*, 385–389.
Akin, A., & Kurbanoglu, I. N. (2011). The relationships between math anxiety, math attitudes, and self-efficacy: A structural equation model. *Studia Psychologica, 53*, 263273.
Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within- and cross-domain relations with achievement and the mediating role of academic self-concept. *Contemporary Educational Psychology, 51*, 240–252.
Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27*(3), 197–205.
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "Choking under pressure" in math. *Psychological Science, 16*(2), 101–105.
Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860–1863.
Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.
Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task-value, and achievement goals. *Journal of Educational Psychology, 93*(1), 23–34.
Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi Block-Tapping Task performance as a function of path configuration. *Journal of Clinical and Experimental Neuropsychology, 27*(1), 127–134.

Cargnelutti, E., Tomasetto, C., & Passolunghi, M. C. (2017). How is anxiety related to math performance in young students? A longitudinal study of grade 2 to grade 3 children. *Cognition & Emotion, 31*(4), 755–764.
Ching, B. H. (2017). Mathematics anxiety and working memory: Longitudinal associations with mathematics performance in Chinese children. *Contemporary Educational Psychology, 51*, 99–113.
Csizér, K., & Kormos, J. (2009). Modeling the role of intercultural contact in the motivation of learning English as a foreign language. *Applied Linguistics, 30*, 166–185.
Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development, 82*(3), 766–779.
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*, e33.
Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology, 7*, 508.
Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*(3), 830–847.
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127.
Fennema, E., & Sherman, J. A. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal, 14*, 51–71.
Ganley, C. M., George, C. E., Cimpian, J. R., & Makowski, M. B. (2018). Gender equity in college majors: Looking beyond the STEM/non-STEM dichotomy for answers regarding female participation. *American Educational Research Journal, 55*(3), 453–487.
Geary, D., Hoard, M. K., Nugent, L., Chu, F. W., Scofield, J. E., & Hibbard, D. F. (2019). Sex differences in mathematics anxiety and attitudes: Concurrent and longitudinal relations to mathematical competence. *Journal of Educational Psychology, 111*(8), 1447–1461.
Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development, 19*(1), 21–46.
Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66*, 153–166.
Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology, 32*(1), 47–56.
Haworth, C. M. A., Harlaar, N., Kovas, Y., Davis, O. P. S., Oliver, B. R., Hayiou-Thomas, M. E., ... Plomin, R. (2007). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics, 10*(4), 554–563.
Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33–46.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.
Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS): Construction, validity, and reliability. *Assessment, 10*(2), 178–182.
Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching, 43*(2), 154–167.
Horwitz, E. K., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125–132.
Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35.
Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398.
IBM Corp (2016). *IBM SPSS statistics for windows, version 24.0.* New York: IBM Corp Released 2016.
Jacobs, J. E. (1991). Influence of gender stereotype on parent and child mathematics attitudes. *Journal of Educational Psychology, 83*, 518–527.
Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software, 56*(6), 1–31.
Jones, M. G., & Wheatley, J. (1990). Gender differences in teacher-student interactions in science classrooms. *Journal of Research in Science Teaching, 27*(9), 861–874.
Justicia-Galiano, M. J., Martín-Puga, M. E., Linares, R., & Pelegrina, S. (2017). Math anxiety and math performance in children: The mediating roles of working memory and math self-concept. *British Journal of Educational Psychology, 87*, 573–589.
Kalaycioğlu, D. B. (2015). The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, The Netherlands, Turkey, and the USA. *Educational Sciences: Theory and Practice, 15*(5), 1391–1401.
Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Science of Learning, 3*, 12.
Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment, 27*(3), 206–225.
Lavasani, M. G., Hehazi, E., & Varzaneh, J. (2011). The predicting model of math anxiety: The role of classroom goal structure, self-regulation and math self-efficacy. *Procedia Social and Behavioral Sciences, 15*, 557–562.
Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*, 355–365.
Lindberg, S., Linkersdörfer, J., Ehm, J.-H., Hasselhorn, M., & Lonnemann, J. (2013). Gender differences in children's math self-concept in the first years of elementary school. *Journal of Education and Learning, 2*(3), 1–8.
Löwe, B., Decker, O., Müller, M. S., Brähler, E., Schellberg, D., Herzog, W., et al. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274.
Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520–540.
MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology, 15*(1), 3–26.
MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*(2), 265–287.
Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology, 93*(3), 543–553.
Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence, 27*, 165–179.
Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60–70.
Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*, 591–606.
Moakler, M., & Kim, M. M. (2014). College major choice in STEM: Revisiting confidence and demographic factors. *The Career Development Quarterly, 62*(2), 128–142.
Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research, 4*(2), 1–11.
Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129–1167.
Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The Reciprocal Internal/External Frame of Reference Model: An integration on models of relations between academic achievement and self-concept. *American Educational Research Journal, 48*(6), 1315–1346.
Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, 21*, 467–487.
Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8^{th} ed.). Los Angeles, CA: Muthén & Muthén.
Namkung, J. M., Peng, P., & Li, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis.

Review of Educational Research, 89(3), 459–496.
Nie, Y., Lau, S., & Liau, A. K. (2011). Role of academic self-efficacy in moderating the relation between task importance and test anxiety. *Learning and Individual Differences, 21*, 736–741.
Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44–59.
Pagulayan, K. F., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the corsi-block-tapping task. *Journal of Clinical and Experimental Neuropsychology, 28*, 1043–1052.
Pekrun, R. (1992). The expectancy-value theory of anxiety: Overview and implications. In D. G. Forgays, T. Sosnowski, & K. Wrzesniewski (Eds.). *Anxiety: Recent developments in self-appraisal, psychophysiological and health research* (pp. 23–41). Washington, DC: Hemisphere.
Pekrun, R. (2006). The Control-Value Theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.
Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*(5), 1653–1670.
Perkun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.). *Emotion in education* (pp. 13–36). Amsterdam: Academic Press.
Ramirez, G., McDonough, I. M., & Jin, L. (2017). Classroom stress promotes motivated forgetting of mathematics knowledge. *Journal of Educational Psychology, 109*(6), 812–825.
Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist, 53*(3), 145–164.
Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's progressive Matrices and vocabulary scales.* Oxford, UK: Oxford University Press.
Rimfeld, K., Malanchini, M., Kropohl, E., Hannigan, L. J., Dale, P. S., & Plomin, R. (2018). The stability of educational achievement across school years is largely explained by genetic factors. *NPJ Science of Learning, 3*, 16.
Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction, 34*, 32–41.
Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*, 241–252.
Smith, E. (2013). Women into science and engineering? Gendered participation in higher education STEM subjects. *British Educational Research Journal, 37*(6), 993–1014.
Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholiguistics, 30*, 725–755.
Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363–374.
Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 20*, 173–180.
Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, A. (2014). Reactive Recruitment of attentional control in math anxiety: An ERP study of numeric conflict monitoring and adaptation. *PloS One, 9*(6), e99579.
Tobias, S. (1986). Anxiety and cognitive processing of instruction. In R. Schwarzer (Ed.). *Self-related cognitions in anxiety and motivation* (pp. 35–54). Hillsdale, NJ: Lawrence Erlbaum Associates.
Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology, 38*(1), 1–10.
Xu, M. K., Marsh, H. W., Hau, K.-T., Ho, I. T., Morin, A. J. S., & Abdujabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. *Journal of Educational Psychology, 105*(2), 489–503.