Running head: EFFECTS EVALUATIVE MESSAGES ON MATHEMATICS

An Evaluative Message Fosters Mathematics Performance in Male Students but Decreases Intrinsic Motivation in Female Students

Angelica Moè[1] & David W. Putwain[2]

[1] Department of General Psychology – University of Padova–Italy

[2] School of Education, Health and Community - Liverpool John Moores University – United Kingdom

Author Notes

Correspondence concerning this article should be addressed to Angelica Moè, via Venezia 8, 35131 Padova, Italy, e-mail angelica.moe@unipd.it, tel. ++39-049-8276689, fax ++39-049-8276600.

Acknowledgments

We are grateful to dr. Mariarosa Bicego for her help with data collection. The present work was carried out within the scope of the research program Dipartimenti di Eccellenza (art.1, commi 314-337 legge 232/2016), which was supported by a grant from MIUR to the Department of General Psychology, University of Padua.

Declaration of interest

The authors declare that they have no conflict of interest.

Abstract

This study contrasted the effects of two task messages, evaluative or non-evaluative, on mathematics performance, affect, and intrinsic task motivation. One hundred-twenty secondary school students aged 17-21 years were delivered one of the two messages, or assigned to a control condition, before completing a mathematics task, measures of message appraisals (challenge and threat), affect (pleasantness, arousal, dominance), and a behavioural indication of intrinsic task motivation. The evaluative message raised performance only in males, while for females both messages decreased intrinsic motivation for the task, probably due to stereotype threat. Implications for future research and educational practices are discussed.

Keywords: stereotype threat; mathematics; gender; intrinsic motivation; affect

An Evaluative Message Fosters Mathematics Performance in Male Students but Decreases Intrinsic Motivation in Female Students

Teachers routinely deliver messages to students aimed at fostering motivation and improving achievement (e.g., Putwain, Symes, & McCaldin, 2017; von der Embse, Schultz, & Draughn, 2015). While research has shown that encouraging messages (e.g., stressing utility, providing positive expectations, suggesting reasons for engaging, sustaining self-determination, self-efficacy beliefs or effort attribution) foster motivation more than threatening messages (e.g., reminding the negative consequences of failing; for a meta-analysis see Lazowski & Hulleman, 2016), teachers commonly use evaluative messages to warn students about the negative consequences of failure (the so-called fear appeals: Putwain & Roberts, 2012). This raises the question if messages focused on evaluation rather than on strengths (such as effort, value, self-efficacy) could be effective in some conditions and for some students.

In this study we will consider the math domain and compare message effectiveness in boys and girls. Due to a common-held gender stereotype ('women are less skilled than men in math and spatial thinking'), girls could experience a stereotype threat, that is feel afraid to demonstrate they underperform boys (e.g., Spencer, Steele, & Quinn, 1999; Lewis & Sekaquaptewa, 2016). Instructions or experimental conditions (e.g. a single woman with two or more men in the testing situation: e.g., Inzlicht & Ben-Zeev, 2000) can elicit this stereotype thus affecting performance. This is not always demonstrated: for instance, recently Finnigan & Corker (2016) and Flore, Mulder, & Wicherts (2019) failed to confirm such effects on performance. Consequently, research has studied the factors which can nullify or emphasize the stereotype threat effects and found that when the task is introduced as an 'evaluative test', it is valued as important (the so-called domain-identification and gender-identification) the size of the effect increase (for meta-analyses and reviews see Flore &

Wicherts, 2015; Maass & Cadinu, 2003; Nguyen & Ryan, 2008; Picho et al., 2013). Instead, when the gender difference is not ascribed to genetic factors (Dar-Nimrod & Heine, 2006), and it is explained that potential anxiety results not from alleged inability, but by the common-held stereotype (Johns, Schmader, & Martens, 2005), the effects nullify. Moreover, a stereotype threat could affect not only performance, but overall interest and positive affect: in the main, male students are more interested and enjoy math-related fields more than female students (Blue & Gann, 2016), with a very small but significant standardized effect size of 0.10 (Froiland & Davison, 2016). This suggests that messages effectiveness could differ between genders, potentially eliciting stereotype threat or lift effects (Walton & Cohen, 2003) respectively in female and male students. Finally, a correlational approach has been adopted in many studies (e.g., Putwain, Symes, & Remedios, 2016; Symes & Putwain, 2016), while there are few experimental studies assessing the effects of different teacher messages (e.g., Putwain & Best, 2011; 2012; Putwain & Pescod, 2018; von der Embse et al., 2015).

Effects of Teacher Messages on Performance

Teachers mostly deliver messages focused on the cost associated with not engaging in study-related behaviours, for instance study regularly, complete homework, pay attention in class – (Putwain et al., 2016; Putwain et al., 2017), instead of messages framed on advantages. This appears to be due to a belief in their effectiveness (Putwain & Roberts, 2012; Putwain, & von der Embse, 2018), while research has shown that encouraging rather than threatening messages (e.g., stressing that ‘you can’ by outlining the importance of effort for success and suggesting rationales for the usefulness of a task) should foster a range of positive outcomes (Ryan & Deci, 2000; 2017; Wigfield & Eccles, 2000). A confirmation of this rationale comes from research showing that prompting usefulness (e.g., Brisson et al., 2017; Gaspard et al., 2015a; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009; Hulleman, Kosovich, Barron, & Daniel, 2017), mastery goals (e.g.,

Hidi & Harackiewicz, 2000) or a growth mindset (e.g., Blackwell, Trzesniewski, & Dweck, 2007; Yeager et al., 2016) raise achievement. Positive effects of performance-goal oriented messages on performance have been found too (e.g., Elliot & Church, 1997; Harackiewicz, Barron, Carter, Lehto, & Elliot, 1997), suggesting that focusing on evaluation is not always ineffective.

In mathematics, the subject we focus in the present study, it is possible that male students will be favoured by the common-held stereotype of males as more math talented than females thus showing a stereotype lift effect (Walton & Cohen, 2003). On the opposite, female students may experience a stereotype threat, that is the fear to underperform due to the same common-held stereotype (Steele, & Aronson, 1995). For instance, Kellow and Jones (2008) found that an evaluative message favoured performance only for the non-stereotyped group, suggesting that for those who are expected to perform poorly due to a common-held stereotype a message emphasizing the evaluative aspect of performance is not effective because it induces stereotype threat.

Effects of Teacher Messages on Motivation, and Affect

Effects of messages have been assessed mostly by considering examination or task performance, while less attention has been devoted to effects on affect (considering the dimensions of pleasure, arousal and dominance; Bradley & Lang, 1994), and on intrinsic task motivation, defined as motivation for the activity (not for external rewards or goals or compliance) moved by expected feelings of pleasure or satisfaction (Deci, 1975).

Previous research results suggest that both pleasure, arousal, dominance and intrinsic motivation could be affected by the messages delivered by teachers. For instance, Froiland and Worrell (2017) found that parental autonomy support, which implies delivering encouraging messages, favour intrinsic life goals and grades. Putwain and Best (2011) found that a fear appeal message increased anxiety, but also that this increased anxiety did not affect

performance, suggesting that emotions driven by the message delivered were not the underlying mechanism. Putwain and Remedios (2014) found that motivation mediated the indirect relations from teacher messages to performance. However, teacher messages in this study were self-reported and not experimentally manipulated. However, in all these studies intrinsic motivation for the task was assessed through self-reports, while, in the present study, to avoid biased responses due to social desirability or compliance with the experimenter, we will measure it behaviourally as done by Deci (1971).

Beyond anxiety, fear appeals have been shown to increase surprise, sadness, anger, puzzlement (Dillard, Plotnick, Godbold, Freimuth, & Edgar, 1996), irritation (Kirscht & Haefner, 1973), tension (LaTour & Pitts, 1989), and disgust, depression and loss of pleasure (Kohn, Goodstadt, Cook, Sheppard, & Chan, 1982). However, these studies, reviewed by Witte and Allen (2000), considered in detail health attitudes rather than educational outcomes. In educational contexts, a few researchers showed that when students value a subject and believe that with effort success is possible teacher messages are associated with greater positive affect, and motivation (Putwain et al., 2017; Putwain et al., 2016). This suggests that messages delivered to students should affect not only performance, but also affect and intrinsic motivation for the task.

Gender Differences in Message Effectiveness

Male students typically consider themselves more skilled then females in mathematics (e.g., Skaalvik & Skaalvik, 2004), and consequently report more enjoyment and less anxiety than females when referring to math domains (Frenzel, Pekrun, & Goetz, 2007). Parents and teachers play a critical role in shaping these ability-related beliefs (e.g., Frenzel, Goetz, Pekrun, & Watt, 2010; Gunderson, Ramirez, Levine, & Beilock, 2012; Upadyaya & Eccles, 2015), contributing to developing the stereotype of males as more skilled than females in math (Keller, 2001), and the belief that math is more useful for males than females

(Fredricks, Hofkens, Wang, Mortenson, & Scott, 2018; Watt et al., 2012). These parental expectations predict subsequent course taking and math achievement (Froiland & Davison, 2016), leading females experiencing a stereotype threat, which could affect performance and message effectiveness.

Since an evaluative message has been found to positively affect performance in those who perceive to be able (e.g., Elliot & Church, 1997; Harackiewicz, Barron, Carter, Lehto, & Elliot, 1997) male students should be favoured more than females by an evaluative message because the common-held stereotype lead them to believe to perform better than females in mathematics. Female students will show decreased performance and intrinsic motivation, since stereotype threat effects refer to a broader range of aspects, including motivation for the task (e.g., Lewis & Sekaquaptewa, 2016).

Fear and Challenge Appraisals Shape the Message Effectiveness

Appraisals are cognitive evaluations, based on perceptions and individual interpretations of environmental events based on (a) personal relevance for one's goals and well-being and (b) perceived capability to be able to face them (Folkman, 2008; Lazarus, 2006, Skinner & Brewer, 2002). The cognitive-appraisal model of the stress process (Folkman & Nathan, 2011; Lazarus, 2006) suggests that there are two stages: primary appraisal based on estimating the importance of the event (personal relevance), and secondary appraisal based on looking for resources and options to face the situation (personal resources), linked in a reciprocal cycle (Putwain & Symes, 2014). A challenge appraisal occurs when a student anticipates success is likely and (s)he perceives him/herself capable to respond to task demands, whereas a threat appraisal occurs when a student anticipates failure because the task requests overweight his/her perceived abilities to successfully perform it (Putwain & Symes, 2016; Symes & Putwain, 2016).

The same message could be appraised either as a challenge or as a threat (Putwain &

Symes, 2014; Putwain et al., 2016; Putwain, Symes & Wilkinson, 2017), and consequently lead to expectations which will affect differently performance (e.g. Putwain, Symes, & Wilkinson, 2017), motivations (e.g., Putwain, Remedios, & Symes, 2015) and emotions (e.g., Durik, Shechter, Noh, Rozek, & Harackiewicz, 2015). For instance, Putwain et al. (2016) found that the same fear appeal could lead either to a challenge or to a threat appraisal, and, as a consequence, favour or impair, respectively, behavioural engagement. Putwain, Symes, and Wilkinson (2017), showed that a challenge appraisal predicted performance through increased behavioural engagement, while a threat appraisal harmed performance by reducing behavioural engagement.

These studies show that the way that messages are appraised could play an important role in determining the message effectiveness; motivation (Putwain & Symes, 2014), engagement (Putwain et al., 2016), and achievement (Putwain & Symes, 2011), are higher following a challenge appraisal and lower following a threat appraisal: "the critical factor in determining relations with antecedents and outcomes is not the message frequency but how it is appraised" (Putwain et al., 2017, p. 1).

Hence, in this study, as in previous ones (e.g., Symes & Putwain, 2016; Putwain et al., 2016; Putwain & Symes, 2016; Putwain, Symes, & Wilkinson, 2017), we assessed appraisal by asking after message delivery how much the messages were perceived in a favourable way, as something which can be effectively performed, namely as a challenge, or in a harming way as something beyond ones' capabilities, that is as a threat.

Aims and Hypothesis

This study aimed at assessing the effects of an evaluative vs. a non-evaluative message on performance in a mathematics task, affect (pleasure, arousal, dominance), and intrinsic motivation for the task, in male and female high school students. The following hypotheses lead the research:

H1: The evaluative message will foster performance, dominance, arousal, pleasantness and motivation in male students due to the common-held stereotype that mathematics is a male domain;

H2. Due to the evaluative context, and the nature of the task which assesses mathematics abilities, females will experience a stereotype threat leading them to decreased performance, and lower levels of intrinsic motivation, dominance, and pleasure, and higher levels of arousal.

Method

Participants

One hundred and twenty Italian students (M_{age} = 18.24 years, *SD*=.76, age range 17-21) attending the final year of a single vocational high school setting out for becoming cook or waiter participated on a voluntary basis. There were 60 males, and 60 females, mostly Caucasian, assigned randomly to one of three conditions: evaluative message, non-evaluative message, or a no message control (40 each, 20 males and 20 females, see Procedure). The power analysis fixed effects using G*power (Faul, Erdfelder, Lang, and Buchner, 2007) showed that considering 6 independent groups 20 participants for each was the right number to detect an effect with $p < .05$, $f = 0.25$, critical $F = 3.92$.

Measures

Values. They were assessed using the instrument developed by Putwain et al. (2015; 2017), adapted from the *Michigan Study of Adolescent Life Transitions* scales (Eccles, O'Neill, & Wigfield, 2005). Items were translated into Italian by a research assistant and then back translated by a native English speaker. Participants had to rate on a 5-point Likert-type scale (anchoring points 1=not at all, 5=very much) attainment value (e.g., 'How important is it to you to get a good grade in mathematics? '), and utility value (e.g., 'how important is mathematics for you, outside the school?'). Two scores were obtained by averaging the three

items referring to each aspect. Data collected using English versions of these scales have shown good factorial validity and internal reliability (Putwain et al., 2015, 2016). Cronbach alphas in the present study were .74, and .82, for attainment value and utility value, respectively.

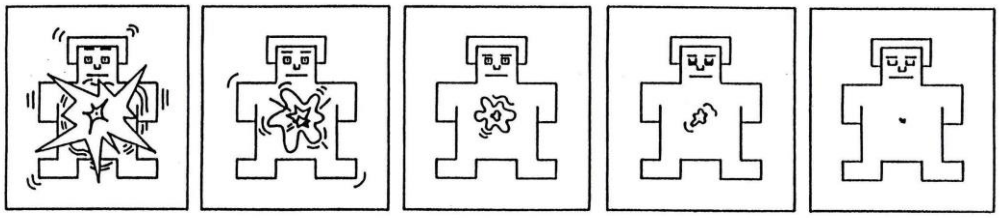
Mathematics performance. It was measured through the AC-FL (Caviola, Gerotto, Lucangeli, & Mammarella, 2016), a math arithmetic test. It consists of 3 sheets each containing 24 operations to perform in 2 minutes. The first sheet contains 24 additions of 2- or 3-digit numbers (e.g., '76 + 103' or '23+3+43'), the second 24 2-digit subtractions (e.g., '69-56' or '72-66') and the third 24 2-digits multiplications (e.g., '45x4' or '37x18'). For scoring the number of correct operations was computed separately for additions, subtractions, and multiplications, as indicated in the manual accompanying the test (Caviola et al., 2016), and done in previous research (e.g., Caviola, Gerotto, & Mammarella, 2016; Caviola, Primi, Chiesi, & Mammarella, 2017; Mammarella, Caviola, Giofrè, & Szűcs, 2018). Cronbach alphas obtained in this study by considering the solved (scored 1) and the unresolved or missed operations (scored 0) were .88, .91, and .80, respectively for additions, subtractions and multiplications respectively. These were very close to those obtained in the validation by Caviola et al. (2016): .89, .90, and .82, for additions, subtractions and multiplications respectively.

Threat and challenge appraisal. The way the messages were appraised was measured using four items, adapted from previous research (Putwain et al., 2015; 2016), two about threat (e.g., 'How worried are you to perform that task?'), two about challenge (e.g., 'How confident are you to do succeed in this task?'). Participants were asked to rate each item by placing a cross along a 16-cm analogue scale. For scoring we calculated two means by averaging the two items regarding challenge appraisal (Cronbach alpha =.70), and threat appraisal (Cronbach alpha =.78).

Affect. The three affective dimensions (pleasantness, arousal, and dominance) were assessed twice, after the message delivery and after the mathematics task, through the Self-Assessment Manikin (SAM: Bradley & Lang, 1994), by asking the participants how they feel right now. It presents schematic figures ranging from a smiling, happy figure to a very unhappy figure (pleasantness), from an excited wide-eyed figure to a sleepy figure with eyes closed (arousal), from a very small figure to a very large dominating the situation (dominance), see Figure 1 for arousal.

Figure 1

Example item of the Self-Assessment Manikin (Bradley & Lang, 1994) for measuring arousal



The task is to put three crosses, one for each dimension (row), on the figure which best represents the affective state, along a continuous nine-point scales. For scoring the single three values assigned by the participants were considered, as done in the validation study (Bradley & Lang, 1994) where each of the affective dimensions were shown to relate with a measure took with a semantic differential method. For a detailed description of the instrument, see Bynion & Feldner (2017). For some recent researches using it see Geethanjali, Adalarasu, Hemapraba, Pravin Kumar, & Rajasekeran (2017), Murdoch, Partin, Vang, & Kehle-Forbes (2019), and Nadler, Cordy, Stengel, Segal, & Hayden (2017). Finally, for better understanding of the results, the scores regarding pleasure and arousal were reversed so that higher scores will mean higher pleasure and arousal.

Intrinsic motivation for the task. It was assessed through a behavioural method based on the procedure proposed by Deci (1971) following which the choice of the task is an index of intrinsic motivation. Participants were told almost of the end of the procedure they had an additional 5 minutes to spend choosing among one of the following three options: a) continue with the mathematics task, b) perform a verbal task (asking, for instance, to write in 2-minutes all the names starting with 'st' which come to mind), c) waiting do nothing. If participants chose to finish the mathematics task, they were required to use a different-colour pencil to differentiate additional mathematics calculations from those performed during the 6-minutes allowed. The choice was coded as follows: 0=do nothing, 1=verbal task, 2=mathematics task.

Procedure

After having obtained the approval from the Departmental Ethical Committee, we contacted three high school principals. One agreed giving the permission for letting the students participate. This done, written parental consent was obtained for participants below the age of 18 years.

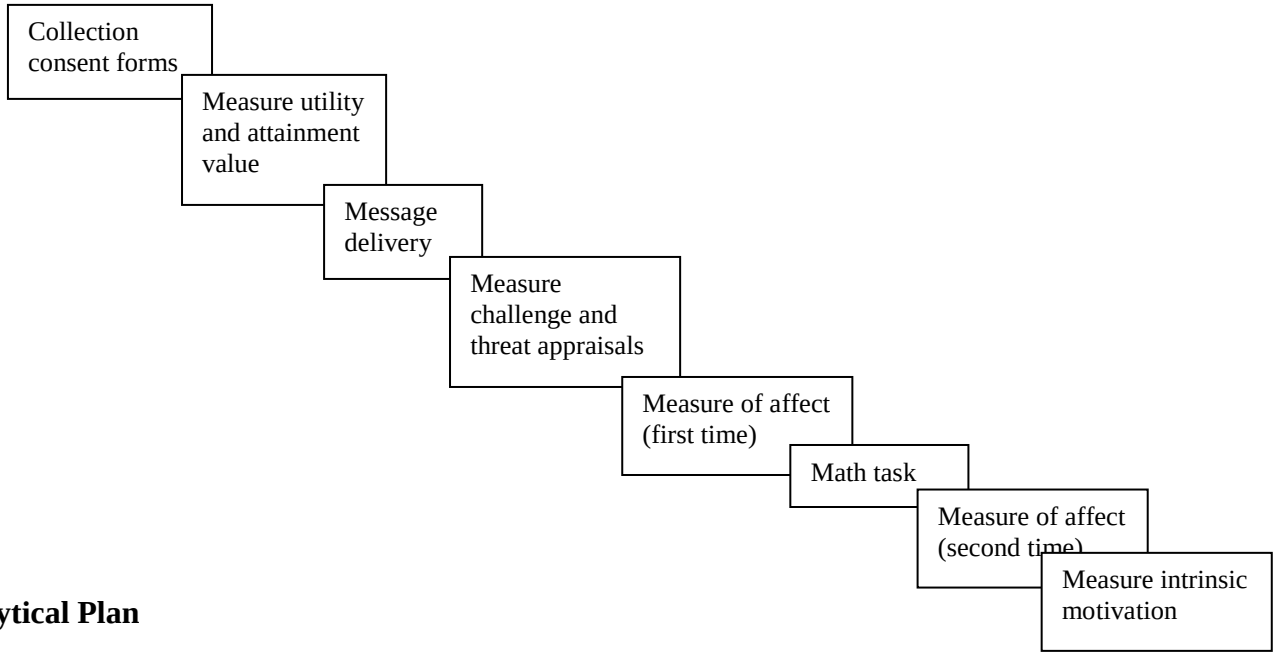
The participants were tested in groups of three in a quiet room in school by an unknown experimenter who introduced herself as a psychologist who will present them scientific evidence. First, they were asked to sign a written consent form, then to complete the utility and attainment value items. This done, they were assigned randomly to one of three conditions: receiving an evaluative message, a non-evaluative message, or no message (control). All the participants were told: ‘The task you are going to perform assess your mathematics abilities, which are very important for your life and future profession’. The evaluative message participants were then delivered the following message: ‘Those **scoring high** can achieve the **highest marks in the final examination** and have **more possibilities to obtain better jobs** or even enter university’ (the words in bold were particularly stressed in

the verbal delivery). The non-evaluative message participants were told ‘In this task **everybody can succeed by putting enough effort,** which is the key for success. Abilities one believes to have do not matter at all. Just do your best and feel confident. Say to yourself: 'I can succeed', and you will perform well’. These messages were read aloud. Participants could also follow them printed on a sheet.

Once delivered the messages participants were asked to answer the threat and challenge appraisal items related to the math task they were going to perform and the three SAM items to assess their affective states in that moment. This done, they were re-read again the messages, asked to perform the mathematics task (2 minutes for each sheet of operations with a 1-minute break between them), to complete the SAM again (second administration, after performing the math task), and finally to choose a final task aimed at measuring their intrinsic motivation towards the math task (see Figure 2 for a graphical representation of the Procedure).

Figure 2

The time-line of the Procedure



Analytical Plan

First descriptive statistics for all the variables were calculated, to verify the mean level of utility and attainment value and overall the mean scores obtained. Second, four between-participants 3 (messages: evaluative, non-evaluative, and control) x 2 (gender) ANOVAs were run to verify there was no difference in pre-message values and appraisals. Second, to test the first proposition of both H1 and H2, effects on mathematics performance were assessed through a series of 3 (message: evaluative, non-evaluative, and control) ANCOVAs, run separately by each gender, respectively on mean number of additions, subtractions, and multiplications correctly solved, with threat and challenge appraisals as covariates. Third, to test the second proposition of both H1 and H2 related to effects on affective dimensions and intrinsic motivation, six 3 (Message) x 2 (Time) ANCOVAs with challenge and threat appraisals as covariates were run on mean pleasantness, arousal, and dominance ratings in males and females. Message (evaluative, non-evaluative, or control) was the between-participants factors. Time (after message delivery vs. after math task) was the within-participants factor. Then two chi-square analyses were run (one for each gender) to examine differences in intrinsic motivation for the task due to the messages delivered in the two genders.

Analyses and Results

Preliminarily Analyses

The descriptive analyses (see Table 1) showed that mean scores regarding values were respectively 3.41 and 2.80 for attainment and utility, in the middle of the range 1 to 5, slightly below than those observed in previous research (e.g., Putwain et al., 2016; Symes & Putwain, 2016). The mean number of correctly solved operations was very close to the values obtained in the validation (16.46, 13.49, 9.72 for additions, subtractions and multiplications respectively: Caviola et al., 2016) and within the validation ranges. The challenge appraisal was more than double of the threat appraisal, suggesting that the situation was perceived

much more challenging than threatening. The affective dimensions were in the middle of the theoretical ranges, suggesting a mild emotional engagement.

Table 1

Descriptive Statistics for Values, Performance, Appraisals, Affect, and Intrinsic Motivation for the Task

Variable	*M*	*SD*	Actual Range (Theoretical)
Values			
Attainment value	3.41	0.75	1.33-5 (1-5)
Utility value	2.80	0.83	1-5 (1-5)
Appraisals			
Challenge appraisal	10.94	3.28	0-16 (0-16)
Threat Appraisal	4.59	3.93	0-13.85 (0-16)
Performance			
Additions correctly solved	17.80	4.44	3-24 (0-24)
Subtractions correctly solved	14.46	5.47	0-24 (0-24)
Multiplications correctly solved	9.13	3.25	0-15 (0-24)
Affect			
Pleasure after message	3.64	0.74	1-5 (1-5)
Arousal after message	2.99	1.00	1-5 (1-5)
Dominance after message	3.48	0.89	1-5 (1-5)
Pleasure after math performance	3.34	0.91	1-5 (1-5)
Arousal after math performance	3.15	0.94	1-5 (1-5)
Dominance after math performance	3.31	0.91	1-5 (1-5)
Intrinsic motivation for the task	1.24	0.53	0-2 (0-2)

Participants assigned to the three conditions did not differ in attainment value, utility value, challenge appraisal, nor threat appraisal (see Table 2 for mean values). There was only one effect due to gender about utility value [$F(1, 114)=6.16$, $p=.029$, $\eta_p^2=.05$; males $M=2.99$, $SD=0.84$; females $M=2.62$, $SD=0.79$)] and no significant interaction. This confirms we could include threat and challenge appraisal as covariates.

Table 2

Mean (standard deviations) in values and appraisals split by gender and condition

Gender	Condition-message	Measure			
		Utility value	Attainment value	Threat appraisal	Challenge appraisal
Males	Evaluative	3.00 (0.72)	3.40 (0.83)	3.71 (4.00)	10.72 (4.71)
	Non-evaluative	3.15 (0.85)	3.32 (0.77)	3.92 (4.17)	11.03 (3.22)
	Control	2.82 (0.93)	3.52 (0.72)	4.35 (3.90)	10.70 (3.64)
Females	Evaluative	2.72 (0.79)	3.42 (0.81)	5.62 (3.45)	10.93 (2.49)
	Non-evaluative	2.52 (0.89)	3.27 (0.86)	6.04 (4.33)	11.20 (2.60)
	Control	2.62 (0.70)	3.57 (0.56)	3.89 (3.53)	11.07 (2.90)

Effects on Mathematics Performance

The ANCOVAs revealed a significant effect message only for males and for additions [$F(2, 55)=3.77$, $p=.015$, $\eta_p^2=.12$] and subtractions [$F(2, 55)=3.61$, $p=.034$, $\eta_p^2=.12$]. The post-hoc analyses showed that males solved correctly more additions [$t(38)=2.11$, $p=.041$, Cohen $d=1.23$] and more subtractions [$t(38)=2.44$, $p=.019$, Cohen $d=1.45$] in the evaluative, compared to the non-evaluative message condition (see Table 3 for mean values). None of the comparisons with control condition were significant at $p<.050$.

Table 3

Mean Operations Solved in the Three Conditions splitted by Gender (Standard Deviation in parentheses).

Gender	Condition-message	Operations		
		Additions	Subtractions	Multiplications
Males	Evaluative	19.20 (4.29)	18.10 (4.69)	9.25 (3.08)
	Non-evaluative	15.85 (5.64)	13.80 (6.33)	8.90 (3.92)
	Control	18.20 (4.49)	14.70 (5.94)	8.35 (3.63)
Females	Evaluative	17.75 (4.00)	13.50 (3.56)	9.30 (2.77)
	Non-evaluative	17.60 (4.60)	13.60 (6.26)	9.30 (3.43)
	Control	18.20 (3.12)	13.05 (4.44)	9.65 (2.74)

For males the effects challenge and threat appraisal were significant for all the three kind of operations: additions, $F(1, 55)=16.92$, $p<.001$, $\eta_p^2=.23$, $F(1, 55)=10.16$, $p=.002$, $\eta_p^2=.16$; subtractions, $F(1, 55)=7.76$, $p=.007$, $\eta_p^2=.12$, $F(1, 55)=7.56$, $p=.008$, $\eta_p^2=.12$; multiplications, $F(1, 55)=10.08$, $p=.002$, $\eta_p^2=.15$, $F(1, 55)=12.78$, $p=.001$, $\eta_p^2=.19$; respectively for challenge and threat appraisals. For females only threat appraisal was significant [$F(1, 55)=4.47$, $p=.039$, $\eta_p^2=.07$] for subtractions. For males, the higher the challenge appraisal, the higher the number of correctly solved additions ($r=.348$, $p=.006$), subtractions ($r=.226$, $p=.083$), and multiplications, ($r=.251$, $p=.053$). The higher the threat appraisal, the lower the number of correctly solved additions ($r=-.219$, $p=.093$), subtractions ($r=-.240$, $p=.064$), and multiplications, ($r=-.322$, $p=.012$). For females, the higher the threat appraisal, the lower the number of correctly solved subtractions ($r=-.238$, $p=.067$).

Effects on Pleasure, Arousal, and Dominance

The ANCOVA on arousal revealed a significant interaction message by time only for males: $F(2, 55)=4.20$, $p=.020$, $\eta_p^2=.13$. Arousal increased only in the evaluative message condition, from $M=2.70$, $SD=1.03$ to $M=3.27$, $SD=1.07$, $t(19)=2.88$, $p=.010$, Cohen $d=0.54$.

Moreover, a significant effect challenge appraisal was found for males in arousal [$F(1, 55)=4.43$, $p=.040$, $\eta_p^2=.07$] , while for females the interaction challenge appraisal x time on arousal was significant, $F(1, 55)=5.33$, $p=.025$, $\eta_p^2=.09$. The higher the challenge appraisal, the higher the self-reported arousal ($r=.282$, $p=.029$) by males. For females, a challenge appraisal related with arousal only after messages delivery ($r=.302$, $p=.019$).

There was a significant effect threat appraisal for females in arousal, $F(1, 55)=5.19$, $p=.027$, $\eta_p^2=.09$, and for males in dominance, $F(1, 55)=8.52$, $p=.005$, $\eta_p^2=.13$. The higher the threat appraisal the lower the dominance in males ($r=-.288$, $p=.026$), and the higher the arousal in females ($r=.293$, $p=.023$).

Effects on Intrinsic Motivation for the Task

Table 3 reports the number of participants choosing to complete the mathematics task (and index of intrinsic motivation) or to perform a verbal task or do nothing, splitted by gender. Since only four males and two females in different conditions choose to do nothing letting two cells empty, we collapsed the choice ‘do nothing’ with that referring to performing a verbal task into a single category ‘other’ which means ‘not choosing to finish the math task’ suggesting a low level of intrinsic motivation for the task.

Table 4

*Number of Participants Choosing to Finish the Mathematics Task (*an *Index of Intrinsic Motivation for the Task), or to Perform a Verbal Task/do Nothing (collapsed into Other)*

	Males		Females	
Condition-message	Math Task	Other	Math Task	Other
Evaluative	7	13	3	17
Non-evaluative	7	13	4	16
Control	4	16	10	10

The chi-squared analysis on the males' frequencies was not significant [$\chi^2(2)=1.429$, ns] showing that they choose to finish the math task to the same extent in the three conditions. For females, instead, the chi-squared analysis was significant [$\chi^2(2)=7.059$, $p=.029$]. As seen in Table 4 they preferred to finish the math task less after both the evaluative and non-evaluative messages than in the control condition, suggesting decreased intrinsic motivation for the task after those messages.

Discussion

This study compared the effects of an evaluative vs. a non-evaluative message on mathematics performance, affective dimensions and intrinsic motivation in male and female students. The evaluative message focused on the importance to perform high, while the non-evaluative stressed the importance of effort put in doing the task.

A common-held stereotype lead people to consider males as more math skilled then females. When this stereotype is raised explicitly (by instructions) or implicitly by the testing situation females tend to underperform while males are unaffected or even improve performance due to a stereotype lift effect (Spencer et al., 1999; Walton & Cohen, 2003). Male and female students could therefore be differently affected by the messages delivered.

In fact, we predicted that an evaluative message will favour performance in male students and that female students will not be favoured by the two messages (due to experiencing stereotype threat). Below we will discuss them, in turn.

Males are Favoured by an Evaluative Message

The results showed that the evaluative message raised performance in comparison with the non-evaluative, but only for males, thus confirming H1, as about effects on performance. The benefit was very large (Cohen's *d*s up to 1) suggesting that an evaluative message raises performance, in comparison with a non-evaluative message, of up to one standard deviation. Interestingly, this effect, as predicted, applied only for those who are expected to be capable on the basis of the common-held stereotype of math as a masculine subject, and hence who can experience a stereotype lift. Interestingly the effect was even higher that that obtained in previous research (for a meta-analysis see Walton & Cohen, 2003) suggesting that emphasizing the consequences of a good performance (evaluative message) can make the stereotype lift effects stronger. Instead, focusing on effort (non-evaluative message), rather than on the consequences of evaluation, did not give rise to a stereotype lift. This result adds to the literature the point that a stereotype lift effect is more akin to occur when the message is framed on performance consequences.

Significant relations were found with challenge and threat appraisals which related positively and negatively with mathematics performance in males. This finding suggests that fostering a challenge appraisal could be beneficial, while a threat appraisal will lead to detrimental effects on performance also with males, who – in fact – showed lower levels of dominance the higher the threat appraisal.

Unexpectedly, there was no difference due to messages on intrinsic motivation in male students. The majority of male students (approximately 2 out of every 3) preferred to do another task: an evaluative message focused on performing well affected performance but did

not increase motivation toward a subsequent similar task. This issue is useful to consider in real class implementations, suggesting that such a message could favour performance in target tasks, but not increase motivation toward future similar tasks.

Females Experience Stereotype Threat thus decreasing Performance, Affect and Motivation for The Task

Performance for females did not differ among conditions, thus partially disconfirming the first proposition of H2: females were not affected by either the evaluative or the non-evaluative messages. However, for subtractions there was a significant effect threat appraisal showing that the higher the threat the lower the females' performance. This suggests that possibly anxiety – which is one of the underlying mechanisms of stereotype threat effects (Maas & Cadinu, 2003) - could have made the evaluative message ineffective for females. In fact, suggesting the importance to score high when the students do not perceive themselves to be competent would raise anxiety (Pekrun, 2006), an emotion which is inversely related with performance (e.g., Raccanello, Brondino, Moè, Stupnisky, & Lichtenfeld, 2018). In this study anxiety was not assessed but could be inferred by the threat appraisal which tended to be higher (even if not significant statistically) after the two messages than in the control condition only for female students.

Females were unresponsive to the evaluative and also the non-evaluative messages, probably as a consequence of the stereotyped nature of mathematics, thus emphasizing they were experiencing a stereotype threat. In fact, previous intervention studies confirmed that when the stereotyped nature of the task is reframed females improve performance, being their expectation to succeed improved (e.g., Dar-Nimrod, & Heine, 2006; Johns, Schmader, & Martens, 2005; Moè, 2016). This study adds to the literature on stereotype threat effects the issue regarding the evaluative (or not) nature of the message, which was not previously considered. The evaluative message reminded the negative consequences of scoring poorly

thus leading to a typical stereotype threat effect (=fear to underperform). The non-evaluative message stressed that with effort everybody could succeed thus rising potentially the threat to demonstrate that nevertheless the efforts put in doing the task the performance is poor thus again resulting in a threat. The results showed that both messages affected negatively the outcomes leading to no increase in performance and a decrease in intrinsic motivation.

Considering effects on the affective dimensions, females showed increased arousal the higher the threat appraisal, while a challenge appraisal related with arousal only at the first time point assessment, but not after performing the mathematics task. Contrary to the hypothesis no effect due to message was found on the affective dimensions. This could probably depend on the fact that participants (males too) expressed low mean levels of pleasantness, arousal and dominance (see Table 1) and lower levels of threat in comparison with challenge (see Tables 1 and 2), showing we tested participants not so motivated to succeed.

As expected, females showed decreased intrinsic motivation preferring a different task, thus showing behaviourally their disengagement and avoidance tendencies, confirming the second proposition of H2. These effects applied for both messages, showing that a non-evaluative message was not so useful, leading to decreased intrinsic motivation for the task in females in comparison with the control condition and to no effect on performance. Among the potential reasons for this lack of effects is that pointing at the importance of effort could not be enough when students show low levels of engagement, due to the occurrence of stereotype threat effects, that is when they fear that, nevertheless the effort put, they will not score high.

Limitations and Future Directions

The results of this study about the differential effects of messages on performance and motivation are interesting. Nevertheless, there are a few limitations which suggests directions

for future research. First, we considered a specific math arithmetic task, based on additions, subtractions, and multiplications and we do not know whether the results here obtained could be generalized to other more complex tasks requiring reasoning or algebra closer to those typically learned in high school. Second, the intrinsic motivation measure is new and behaviourally based which is a strength. The verbal task, however, was chosen by a large majority of participants possibly due to its novelty. Future studies could consider a different behavioural measure which did not contrast a new-already done task. Moreover, also a self-reported measure should be collected, to compare it with the behavioural one. Then, future research could aim increasing intrinsic motivation for the task via autonomy supportive messages (e.g., Froiland, Davison, & Worrell, 2016) or having peers interested in math sharing their math interest (e.g., Bissell-Havran & Loken, 2009). Third, the messages were delivered by the same person, but the instructor was not the teacher, and this could have shaped the effects of messages. Introducing a manipulation check in future research could be helpful to ascertain to what extent the participants believed to the messages. Forth, participants were from only a single vocational school and this requires caution in generalizing the results to different contexts. Future studies in different high schools should be run to confirm the results here obtained. Additionally, it could be advisable to add a measure of mathematics proficiency before running the study to verify that the three groups do not differ in their math level. Fifth, a huge amount of research focused on math anxiety as a factor affecting math motivation, engagement and achievement, related with task avoidance and stereotype threat (for a review see Chang & Beilock, 2016; Maloney & Beilock, 2012; Ramirez, Shaw, & Maloney, 2018). Math anxiety, typically higher in girls and in poor achievers, was not measured in this study. This is a limitation which future research could consider by assessing it and the role played in mediating the effects of the messages delivered. Sixth, since threat and challenge appraisal measures were taken after message

delivery, having included them as covariates could have underestimated the effects as pointed out by Rohrer (2018). Moreover, the appraisals could be measured before message delivery, and using more items. Seventh, the message 'Those scoring high can achieve the highest marks …' could have not been so convincing thus inflating the effects. Moreover, it could have suggested an extrinsic rather than intrinsic motivation to do the task, based on social comparison, which is usually detrimental (Ames, 1992). Future research should consider improving the non-evaluative message by turning 'everybody can succeed by putting enough effort', into a growth mindset language, such as 'your effort will help you to make your mind stronger' (e.g., Dweck, 2015). Finally, our participants were a low number of high school students and from a single school, and we do not know whether the same results will apply with younger students and in other contexts. Finally, effects were obtained in experimental sessions and we do not know if they would apply to more naturalistic settings.

Conclusions

Previous research found that males and females differ in mathematics interests and values (e.g., Wang, 2012) and that these different beliefs should be taken into account in devising interventions to foster motivation (e.g. Gaspard, et al., 2015b; Hulleman et al., 2010). The present study suggests the importance to take into consideration gender and the occurrence of stereotype threat (or lift) effects. An evaluative message could be helpful for males because it raises performance, whilst not affecting intrinsic motivation for the task. However, the same message could not be so useful for females. In fact, it did not affect performance and reduced intrinsic motivation for the task. When a stereotype is in the air, messages focused on increasing emotional rather than cognitive engagement (e.g. fostering pleasantness, for example by saying 'This is a very interesting task that most students like') would be expected to be more effective for improving females' performance and motivation.

References

Bissell-Havran, J. M., & Loken, E. (2009). The role of friends in early adolescents' academic self-competence and intrinsic value for math and English. *Journal of Youth and Adolescence*, *38*(1), 41–50.

Bynion, T. M., & Feldner, M. T. (2017). Self-Assessment Manikin. In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 58-74). New York, NY: Springer.

Blackwell, L., Trzesniewski, K., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Blue, J., & Gann, D. (2016). When do girls lose interest in math and science? *Science Scope, 32*, 44-47.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry*, *25*(1), 49-59.

Brisson, B. M., Dicke, A. L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal*, *54*(6), 1048-1078.

Caviola, S., Gerotto, G., Lucangeli, D., & Mammarella, I. C. (2016). *AC-FL. Prove di fluenza nelle abilità di calcolo per il secondo ciclo della scuola primaria* [AC-FL. Fluency in calculation test for the second cycle of primary school]. Trento: Erickson.

Caviola, S., Gerotto, G., & Mammarella, I. C. (2016). Computer-based training for improving mental calculation in third-and fifth-graders. *Acta Psychologica*, *171*, 118-127.

Caviola, S., Primi, C., Chiesi, F., & Mammarella, I. C. (2017). Psychometric properties of the

Abbreviated Math Anxiety Scale (AMAS) in Italian primary school children. *Learning and Individual Differences*, *55*, 174-182.

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: A review of current behavioral and psychophysiological research. *Current Opinion in Behavioral Sciences*, *10*, 33-38.

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, *314* (5798), 435.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105-115.

Deci, E. L. (1975). Conceptualizations of intrinsic motivation. In E. L. Deci (Ed.), *Intrinsic motivation* (pp. 23-63). Boston, MA: Springer.

Dillard, J. P., Plotnick, C. A., Godbold, L. C., Freimuth, V. S., & Edgar, T. (1996). The multiple affective outcomes of AIDS PSAs: Fear appeals do more than scare people. *Communication Research*, *23*(1), 44-72.

Durik, A. M., Shechter, O. G., Noh, M., Rozek, C. S., & Harackiewicz, J. M. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion*, *39*(1), 104-118.

Dweck, C. (2015). Carol Dweck revisits the growth mindset. *Education Week*, *35*(5), 20-24.

Eccles, J. S., O'Neill, S. A., & Wigfield, A. (2005). Ability self-perceptions and subjective task-values in adolescents and children. In K.A. Moore & L.H. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 239-247). New York, NY: Springer.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality, 63,* 36-43.

Flore, P. C., Mulder, J., & Wicherts, J. M. (2019). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: a registered report. *Comprehensive Results in Social Psychology, 3*(2), 140-174.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*(1), 25-44.

Folkman, S. (2008). The case for positive emotions in the stress process. *Anxiety, Stress and Coping*, *21*(1), 3–14.

Folkman, S., & Nathan, P. E. (Eds.). (2011). *The Oxford handbook of stress, health, and coping*. Oxford: Oxford University Press.

Fredricks, J. A., Hofkens, T., Wang, M. T., Mortenson, E., & Scott, P. (2018). Supporting girls' and boys' engagement in math and science learning: A mixed methods study. *Journal of Research in Science Teaching, 55*(2), 271-298.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507-537.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*(4), 497-514.

Froiland, J. M., & Davison, M. L. (2016). The longitudinal influences of peers, parents, motivation, and mathematics course-taking on high school math achievement. *Learning and Individual Differences, 50,* 252-259.

Froiland, J. M., Davison, M. L., & Worrell, F. C. (2016). Aloha teachers: teacher autonomy support promotes Native Hawaiian and Pacific Islander students' motivation, school belonging, course-taking and math achievement. *Social Psychology of Education*, *19*(4), 879-894.

Froiland, J. M., & Worrell, F. C. (2017). Parental autonomy support, community feeling and student expectations as contributors to later achievement among adolescents. *Educational Psychology*, *37*(3), 261-271.

Gaspard, H., Dicke, A. L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015a). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226-1240.

Gaspard, H., Dicke, A. L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015b). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, *107*(3), 663-677.

Geethanjali, B., Adalarasu, K., Hemapraba, A., Pravin Kumar, S., & Rajasekeran, R. (2017). Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Research*, *28*, 18-24.

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*(3-4), 153-166.

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, *73*(6), 1284-1295.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical

issue for the 21st century. *Review of Educational Research, 70*(2), 151-179.

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*(4), 880-895.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326* (5958), 1410–1412.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*(5), 365-371.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle. Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*(3), 175–179.

Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology, 141*(2), 165-173.

Kellow, J. T., & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. *Journal of Black Psychology, 34*(1), 94-120

Kirscht, J. P., & Haefner, D. P. (1973). Effects of repeated threatening health communications. *International Journal of Health Education, 16*, 268-277.

Kohn, P. M., Goodstadt, M. S., Cook, G. M., Sheppard, M., & Chan, G. (1982). Ineffectiveness of threat appeals about drinking and driving. *Accident Analysis & Prevention, 14*(6), 457-464.

LaTour, M. S., & Pitts, R. E. (1989). Using fear appeals in advertising for AIDS prevention in the college-age population. *Journal of Health Care Marketing, 9*(3), 5-14.

Lazarus, R. S. (2006). *Stress and emotion*. New York, NY: Springer.

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research, 86*(2), 602–640.

Lewis Jr, N. A., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology*, *11*, 40-43.

Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. *European Review of Social Psychology*, *14*(1), 243-275.

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404-406.

Mammarella, I. C., Caviola, S., Giofrè, D., & Szűcs, D. (2018). The underlying structure of visuospatial working memory in children with mathematical learning disability. *British Journal of Developmental Psychology*, *36*(2), 220-235.

Moè, A. (2016). Teaching motivation and strategies to improve mental rotation abilities. *Intelligence*, *59*, 16-23.

Murdoch, M., Partin, M. R., Vang, D., & Kehle-Forbes, S. M. (2019). The psychological risk of minimal risk activities: A pre-and posttest study using the Self-Assessment Manikin. *Journal of Empirical Research on Human Research Ethics*, *14*(1), 15-22.

Nadler, R., Cordy, M., Stengel, J., Segal, Z. V., & Hayden, E. P. (2017). A brief mindfulness practice increases self-reported calmness in young children: a pilot study. *Mindfulness*, *8*(4), 1088-1095.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*(6), 1314-1334.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315-341.

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology*, *153*(3), 299-333.

Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: Effects on test anxiety and test grade. *Learning and Individual Differences*, *21*(5), 580-584.

Putwain, D. W., & Best, N. (2012). Do highly test anxious pupils respond differentially to fear appeals made prior to a test? *Research in Education*, *88*(1), 1–10.

Putwain, D. W., Nakhla, G., Liversidge, A., Nicholson, L. J., Porter, B., & Reece, M. (2017). Teachers use of fear appeals prior to a high-stakes examination: Is frequency linked to perceived student engagement and how do students respond? *Teaching and Teacher Education, 61,* 73–83.

Putwain, D. W., & von der Embse, N. P. (2018). Teachers use of fear appeals and timing reminders prior to high-stakes examinations: pressure from above, below, and within. *Social Psychology of Education*, *21*(5), 1001-1019.

Putwain, D. W., Nicholson, L. J., Nakhla, G., Reece, M., Porter, B., & Liversidge, A. (2016). Fear appeals prior to a high-stakes examination can have a positive or negative impact on engagement depending on how the message is appraised. *Contemporary Educational Psychology*, *44-45*, 21-31.

Putwain, D., & Remedios, R. (2014). The scare tactic: Do fear appeals predict motivation and exam scores? *School Psychology Quarterly*, *29*(4), 503-516.

Putwain, D. W., Remedios, R., & Symes, W. (2015). Experiencing fear appeals as a challenge or a threat influences attainment value and academic self-efficacy. *Learning and Instruction, 40*, 21–28.

Putwain, D. W., & Roberts, C. M. (2012). Fear and efficacy appeals in the classroom: The secondary teachers' perspective. *Educational Psychology, 32*(3), 355–372

Putwain, D., & Symes, W. (2011). Perceived fear appeals and examination performance: Facilitating or debilitating outcomes? *Learning and Individual Differences, 21*(2), 227-232.

Putwain, D. W., & Symes W. (2014). The perceived value of maths and academic self-efficacy in the appraisal of fear appeals used prior to a high-stakes test as threatening or challenging. *Social Psychology of Education, 17*(2), 229-248.

Putwain, D. W., & Symes, W. (2016). Expectancy of success, subjective task-value, and message frame in the appraisal of value-promoting messages made prior to a high-stakes examination. *Social Psychology of Education, 19*(2), 325-343.

Putwain, D. W., Symes, W., & McCaldin, T. (2017). Teacher use of loss-focused, utility value messages, prior to high-stakes examinations, and their appraisal by students. *Journal of Psychoeducational Assessment, 37*(2), 169-180.

Putwain, D. W., Symes, W., & Remedios, R. (2016). The impact of fear appeals on subjective-task value and academic self-efficacy: The role of appraisal. *Learning and Individual Differences, 51*, 307-313.

Putwain, D. W., Symes, W., & Wilkinson, H. M. (2017). Fear appeals, engagement, and examination performance: The role of challenge and threat appraisals. *British Journal of Educational Psychology, 87*(1), 16-31.

Raccanello, D., Brondino, M., Moè, A., Stupnisky, R., & Lichtenfeld, S. (2018). Enjoyment, boredom, anxiety in elementary schools in two domains: Relations with achievement. *The Journal of Experimental Education, 87*(3), 449-469.

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.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal

models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27-42.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York: Guilford Press.

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*(3-4), 241–252

Skinner, N., & Brewer, M. (2002). The dynamics of threat and challenge appraisals prior to stressful achievement events. *Journal of Personality and Social Psychology, 83*(3), 678-692.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Symes, W, & Putwain, D. W. (2016). The role of attainment value, academic self-efficacy, and message frame in the appraisal of value-promoting messages. *British Journal of Educational Psychology, 86*(3), 446-460.

Upadyaya, K., & Eccles, J. (2015). Do teachers' perceptions of children's math and reading related ability and effort predict children's self-concept of ability in math and reading? *Educational Psychology, 35*(1), 110-127.

von der Embse, N. P., Schultz, B. K., & Draughn, J. D. (2015). Readying students to test: The influence of fear and efficacy appeals on anxiety and test performance. *School Psychology International, 36*(6), 620-637.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social*

Psychology, *39*(5), 456-467.

Wang, M. T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology*, *48*(6), 1643-1657.

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology, 48*(6), 1594-1611.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior*, *27*(5), 591-615.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... Dweck, C.S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374-391.