## Math and Metacognition: Resolving the Paradox

# Shanna Erickson (serickson@ucmerced.edu) Evan Heit (eheit@ucmerced.edu)

Cognitive and Information Sciences, University of California, Merced Merced, CA 95343 USA

#### Abstract

Metacognition plays a fundamental role in forming selfjudgments of ability and knowledge. Is metacognition domain and gender specific? Metacognitive judgments and performance were measured across biology, literature, and math content. Undergraduates took three shortened SAT II Subject Tests, and provided estimates of their performance both before and after taking each test. The results were that differed across judgments domain and gender. Overconfidence was evident in all domains, although estimates of ability were more accurate after taking a test. Males tended to be more overconfident, while females were less confident yet more accurately calibrated when estimating ability. Students were over-confident in math, bringing into question the existence of math phobia. Improvement in calibration and gender difference in calibration were most noticeable in math.

**Keywords:** metacognition, math anxiety, gender differences, mathematics education.

## Introduction

Metacognition, a form of higher-order thinking, plays an important role in cognitive processing. It impacts several areas within cognitive science, such as attention, memory, perception, comprehension, and problem solving (Kitchener, 1983; Metcalfe & Shimamura, 1994). Metacognition aids intellectual endeavors requiring complex thought processes (Schoenfeld, 1983) and also affects social behavior (Jaccard, Dodge & Guilamo-Ramos, 2005) and decision making (Cohen, Freeman, & Thompson, 1998).

Two components of metacognition are of particular importance in education: the ability to monitor what you know, which acts as a basis for predicting retention, and the control processes that are used to enact study activities (Nelson & Dunlosky, 1991). Students need to use metacognitive control in gauging what they know and deciding what study methods to use (Thiede, Anderson, & Therriault, 2003; Metcalfe, 2009). This process is constantly changing, as students adapt their behaviors in monitoring a learning goal. Self-regulation is necessary for this process (Kornell & Bjork, 2007), thus students must select from a variety of strategies, enacting these strategies in goal-directed activities, and monitoring their progress in using these strategies.

Success of metacognition affects students' academic performance (Hattie, Biggs, & Purdie, 1996; Paris & Paris, 2001, Coutinho, 2008), as well as their ability to communicate what they know about a particular problem. Being able to communicate their level of understanding to instructors is crucial to the learning process. It guides how classroom and self-study materials are constructed, and can

affect what strategies students are taught for study and examination.

Metacognition has been shown to play a crucial role in gauging one's own knowledge (Sperling, Howard, Stanley & DuBois, 2004; Schunk & Ertmer, 2000), including specific academic domains such as reading comprehension (Pressley, 2002), math (Pugalee, 2001), science (Schraw, Crippen, & Hartley, 2006), and writing (Pugalee, 2001). Any improvements in metacognition would allow students to better judge what they know and how well they will be able to recall information. This holds much promise for improving student academic performance.

Despite the importance of metacognition, people commonly display glaring overconfidence in their selfperception of their own knowledge and various abilities (Kruger & Dunning, 1999; Dunning, Johnson, Ehrlinger, & Kruger, 2003). Furthermore, people with lower abilities show an even more exaggerated overconfidence. Students in particular often self-report confidence judgments that are unrelated to their actual performance on assessments (Schraw, 1996). Compounding this is students' inability to allocate study times effectively. Methods of self-guided study often result in non-optimal allocation of study time (Son & Sethi, 2010). Improved methods are available, but students generally do not employ them, even though it has been shown that it is possible to use metacognitive control. There is potential for optimal study (Son & Sethi, 2006), but students instead use uninformed metacognitive decisions to structure their study time.

A possible exception to the overconfidence phenomenon is the occurrence of math anxiety (Meece, Wigfield, & Eccles, 1990; Furner & Berman, 2003). Math anxiety (or math phobia) is a fear of math that leads to math avoidance or lower math performance (Ashcraft, 2002; Ashcraft & Krause, 2007) and has been observed in children and adults alike (Wigfield & Meece, 1988). This sometimes extreme anxiety is harmful in educational and workplace settings (Meece et al., 1990; Furner & Berman, 2003), undermining national and worldwide priorities to emphasize science, technology, engineering, and math (STEM) achievement. Indeed, a recent national report predicts increased demand for STEM professionals in the US as well as an inadequate supply of prepared graduates (STEMconnector, 2013). Performing math tasks in stressful situations, such as during tests, only compounds math anxiety (Beilock, 2008). Mathphobic attitudes of teachers can also be detrimental to students' math achievement, particularly for female teachers and students (Beilock, 2010).

This fear of math implies that there should be a corresponding underconfidence in self-evaluation of

mathematical ability. The consistent and persisting documentation of widespread math phobia contradicts the finding that people are generally overconfident. How then can we resolve this paradox? We wish to determine if students are as overconfident in math as they are in other academic domains, or if is math an exception to an otherwise global overconfidence.

Past findings indicate that females generally lag behind their male counterparts on standardized test performance in math (Brown & Josephs, 1999). This is particularly true among high school and college students (Hyde et al., 2006). This gap does appear to have narrowed in recent years (Else-Quest, Hyde, Shibley, Marcia, 2010). However, attitudes toward math between genders still follow differing patterns, and females are more likely to feel intimidated by math than are males (Jakobsson, 2012; Brown & Josephs, 1999). This lack of confidence often leads to a self-fulfilling lag in performance (Brown & Josephs, 1999; Kiefer & Sekaquaptewa, 2007) that can lead to gaps in performance between genders.

We wish to explore is if overconfidence generalizes to all domains of academic knowledge and ability, or if it is If there exist confidence differences domain specific. among various academic subjects, this suggests that overconfidence is domain specific and not a general phenomenon that is implied by the findings of Kruger & Dunning (1999) and Dunning et al. (2003). overconfidence is a global phenomenon, we would expect to see overconfidence in students' ratings across various academic domains. If metacognition is instead domain specific, we would then expect to find differences in overconfidence among academic domains. In the light of math phobia, we would expect to see underconfidence rather than overconfidence in math tasks, in contrast to other domains.

We also seek to determine if metacognitive ability differs over gender as well, keeping in mind that female students show greater math phobia than males. Finally, we compared metacognitive judgments before and after an intervention, namely taking a test, to determine if students are able to improve their metacognitive judgments. We expected to see improvements, as people could re-evaluate their metacognitive estimates after being exposed to more information in the intervention. This would be consistent with Bayesian accounts of cognition, in that people would be updating their hypothesis of ability based on new observations (Jones & Love, 2011; Heit & Erickson, 2011).

## Experiment

We considered test performance, confidence, and calibration in predicted scores. Three comparisons will be highlighted. The first is the comparison among the three different SAT II Subject Tests to assess if overconfidence is a domain specific or general phenomenon. While predictions (estimates before taking an assessment) provide a measure of general confidence within a subject, postdictions (estimates after taking an assessment) provide a more

accurate and comparable measure of metacognitive ability to evaluate knowledge. The use of SAT II Subject Test sample questions gave participants a reference for difficulty level of the assessment before they take it. However, it might have been some years since the participants have taken these, and some participants may have chosen to take a different selection of subject tests than the ones presented in this experiment. Use of retrieval fluency and recognition heuristics would negatively affect metacognition, both for past experience and future performance (Benjamin, Bjork, & Hirshman, 1998). The use of postdictions brings all participants to a more equitable level of familiarity with the test material before making a judgment of ability. As such, calibration was determined by comparing postdicted estimates of performance with actual scores of performance. If overconfidence is domain specific, we then expect that metacognitive performance would differ among different domains, and that there would be a higher rate of underconfidence within math. If metacognition is domain general, then a similar level of overconfidence should be observed across all three assessments.

The second comparison will be one made between genders. Males were expected to show higher confidence ratings in math than females. Third is the comparison between predictions and postdictions for performance on a task. This allows us to determine if participants improved their metacognitive judgments after completing a task. We expect that postdictions for performance on a task will be more accurately calibrated than predictions for the same task, and results reflected this.

### Method

**Participants** There were 31 participants in this experiment: 17 female and 14 male. All were UC Merced undergraduates (mean age = 19.03, SD = 0.98) who took the experiment as a form of extra credit in one of their introductory Psychology or Cognitive Science classes.

Tasks and Materials Participants took three tests: a biology, literature, and math test, each consisting of 15 questions. Participants were told before the experiment that they would be taking tests based on SAT II Subject Tests content. Before each test, participants were asked to provide a predicted score (out of 15) for how well they would do. After taking each test, participants provided a postdicted score for how well they thought they performed. They were not told their actual scores on tests.

## Results

Key descriptive results for all participants are shown in Figure 1. The leftmost bar for each category represents average predicted score, the middle bar represents average actual test score, and the rightmost bar represents average postdicted score. Average performance across all tests was 40%. Participants showed general overconfidence in predicted scores before each test. Overconfidence generally persisted in postdicted scores, although drastic reductions in

residual magnitudes show that participants were better able to assess their ability after each test, providing evidence for improvement in metacognitive judgment of ability (t = 3.30, df = 60, p < 0.0001). The only test in which participants showed slight underconfidence was biology.

Notably, participants showed high overconfidence in math. The residual for average predicted math score was 35%. This was higher than the residuals for both biology (2.6%) and literature (19%).

Results by gender are shown in Figures 2 and 3. Notable was the difference in calibration between genders. Overall, females were more accurate in self-estimates of ability. Differences between their predictions and scores averaged 11%, compared to 29% for males. Similarly for postdictions, females misestimated their performance by an average of 5% while males misestimated by an average of 14%. Males were generally overconfident both before and after taking each assessment. Overall, females had lower measures of overconfidence. Within literature and math, females began with overestimates of their ability, but their postdictions were more calibrated. Within biology, females actually started underconfident and became even more so after taking this assessment.

Both genders show little trace of math phobia, as shown by their predominant overestimates of performance. Average prediction and postdiction residuals in math were 27% (overconfident) and –5% (underconfident) for females and were 42% and 15% for males. Though participants were generally overconfident with their predictions, they were able to improve their metacognitive judgment accuracy significantly in this domain. Males showed the most marked improvement in calibration in math, and females actually changed their judgments from being overconfident to predominantly underconfident.

In a three-way, predicted versus actual score × academic subject (biology or literature or math) x gender (male or female) ANOVA, there was a main effect of gender F(1,29)= 4.48, MSE = 13.69,  $\eta^2 = 0.08$ , p < 0.05. There was also a significant main effect of predicted (mean = 59.00) versus actual (mean = 40.22) score, F(1,29) = 36.61, MSE = 10.1,  $\eta^2 = 0.50$ , p < 0.0001, indicating overconfidence in predictions. There was also a significant main effect of academic subject, F(2, 116) = 6.59, MSE = 5.33,  $\eta^2 = 0.08$ , p < 0.01. Notice that scores were lowest overall in math. There was a significant interaction between these two variables, F(2, 116) = 16.80, MSE = 5.33,  $\eta^2 = 0.20$ , p < 0.200.0001, indicating that degree of overconfidence depended on academic subject. Overconfidence was greatest in math (predicted score = 62.15, actual score = 27.53). We are careful not to over-interpret the interaction, as actual scores also varied by academic subject. There was also a significant interaction between gender and predicted versus actual score, F(1,29) = 8.19, MSE = 10.1,  $\eta^2 = 0.11$ , p < 0.110.01, providing further evidence that overconfidence depended on gender. The remaining main effects and interactions were not significant.

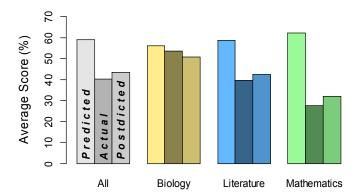


Figure 1: Overall results by test.

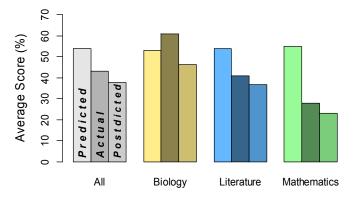


Figure 2: Overall results for females.

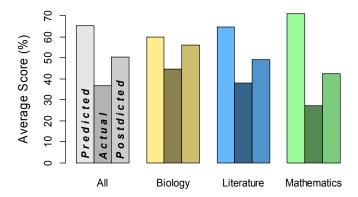


Figure 3: Overall results for males.

We also conducted a similar analysis on postdicted scores (mean = 43.44) and actual scores. This ANOVA revealed a main effect of academic subject, F(2,116) = 36.28, MSE = 4.85,  $\eta^2 = 0.37$ , p < 0.0001, as well as a main effect of gender F(1, 29) = 4.48, MSE = 15.49,  $\eta^2 = 0.09$ , p < 0.05. There was also a significant interaction between gender and postdicted versus actual scores, F(1, 106.27) = 17.55, MSE = 6.05,  $\eta^2 = 0.37$ , p < 0.001, again showing gender differences in overconfidence. Remaining main effects and interactions were not significant.

Note that although scores were lowest for math, this assessment was not designed to be more difficult than the

other subject tests. In fact, it had the lowest difficulty level. During pilot experiments, test questions were chosen using difficulty ratings provided by College Board. Although we originally chose a variety of easy, medium, and difficult questions for each subject test, performance on this balanced math test was so poor that we substituted easier questions in place of all medium and difficult questions. Thus, the severe overconfidence observed in math is not a result of higher test difficulty level compared to other academic subjects.

Figures 4 and 5 show calibration slopes by domain. The dashed line represents the equation y = x (predicted score = actual score) is used to convey perfect calibration. The closer a line is to this dashed line, the better the calibration. For the predicted scores, there are apparent subject differences, e.g., the slope is highest for math, indicating the highest level of sensitivity to actual performance, and the slope is actually slightly negative for biology. Each domain slope more closely follows the calibration line y = x for postdicted scores, showing that that participants were better able to judge their ability after taking each assessment. Each of models also crosses y = x, switching from overconfidence to underconfidence as test performance increases. In addition, correlations between actual scores and residuals calculated from estimated scores (r = -0.63for both predicted and postdicted residuals) reveal that

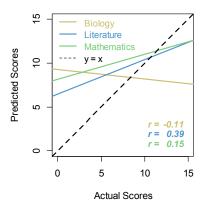


Figure 4: Overall results for postdicted scores.

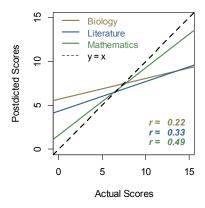


Figure 5: Overall results for females.

higher scores are associated with lower residuals. These findings support the previous work by Kruger & Dunning (1999), Dunning et al. (2003) and Schraw (1996) and show that people with low test scores generally exhibit overconfidence, while people with high test scores are better able to judge their ability. Thus higher performing students tend to be better judges of their ability than are lower performers.

Gender differences were most striking within math. Figure 6 and 7 show calibration models by gender for both predictions and postdictions. Males made predictions with almost no calibration (r = 0.02), and females were overconfident overall with the predictions. Despite this, both genders were able to make much more accurate postdictions. In fact, these postdiction models were the best of any of the observed estimates of ability in this experiment when compared to other subjects.

## **Discussion**

The results of this study support past findings that people are generally overconfident in their abilities, although overconfidence does not appear to be exactly the same across domains. This was shown in the varying judgments of ability across academic domains. While both males and females are generally overconfident, females tend to be better calibrated in judging their domain knowledge.

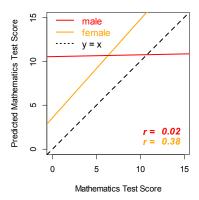


Figure 6: Overall results for females.

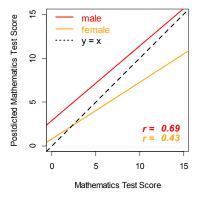


Figure 7: Overall results for females.

Postdictions were significantly lower than predictions, showing that people are able to recalibrate their metacognitive judgments towards more accurate judgments after attempting an assessment.

In addressing the paradox of general overconfidence alongside the seeming exception of math phobia, we saw that overconfidence was particularly high in predicted scores for math assessment. This led us to question whether math phobia was present.

Although there was a higher incidence of overconfidence in mathematics, participants showed the greatest beneficial adjustment of metacognitive judgment miscalibration for mathematical ability. All participants were successfully able to recalibrate their estimates towards more accurate judgments of domain knowledge after an assessment.

We have replicated this severe overconfidence in math in other experiments, although gender no longer reached the level of statistical significance. Thus high math overconfidence is not specific to college-level students: In a subsequent experiment (n = 40), this result was replicated at a local high school using the same experimental design. In another experiment with college students (n = 46), we extended our findings by using the same experiment presented here, but also including Likert scale measures of confidence for each domain, as well as an adapted math Anxiety Rating Scale (MARS) survey (Alexander & Martray, 1989). Initial results indicate that there does exist math phobia, as we observed MARS ratings similar to other college populations identified as math anxious. expected higher anxiety ratings to be linked underestimates of ability. We did not observe this. Instead, math phobia moderated overestimates of ability to be less extreme, although overconfidence still persisted. This is a possible explanation for the coexistence of math phobia and overconfidence in mathematical ability. Further plans include the replication of studies within actual classroom settings in which participants must judge their ability on class assessments.

Does a higher confidence in one subject over another really indicate domain specificity rather than generality? If so, this suggests we may be using different metacognitive methods for different domains such as sciences versus the humanities. Alternatively, we might be using one overarching metacognitive ability that uses different cues and leads to different results across domains.

Our results are relevant for applications in cognitive science, particularly for studying and improving education. We have seen that students are overconfident in math, yet there is evidence that these same students are math phobic. These views pose two strong deterrents for students to seek practice and improvement in math. If students are overconfident in their mathematical abilities and have anxiety about mathematical tasks, they have little incentive to study the subject. This reluctance likely carries over to other science, technology, and engineering subjects that require a significant amount of math background.

We also know that our use of metacognition does not always lead to calibrated self-views of ability. There are optimal models for allocation of study time, but student behaviors do not conform to these (Son & Sethi, 2006; Judgments of improvement and learning rate that students use to make time allocation decision are often inaccurate as well (Townsend & Heit, 2010; 2011). Math phobia has come to be so expected that it has started to influence curriculum design. Already changes have been made in computer science programs to deemphasize math (Tucker, 2001) even though math content is fundamental to The spread of this trend to other science, this area. technology and engineering programs would seriously undermine students' foundational math knowledge. Therefore it will be important to develop techniques that improve students' metacognitive calibration mathematics and other subjects.

## References

- Alexander, L., & Martray, C. R. (1989). The development of an abbreviated version of the math Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development*.
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11(5), 181-185.
- Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243-248.
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, 17(5), 339-343.
- 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.
- Benjamin, A. S., Bjork, R. A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterization of the utility of subjective fluency. *Acta Psychologica*, 98(2), 267-290.
- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of personality and social psychology*, 76(2), 246.
- Cohen, M. S., Freeman, J. T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. *Making decisions under* stress: Implications for individual and team training, 155-190.
- 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.
- Coutinho, S. (2008). Self-efficacy, metacognition, and performance. *North American Journal of Psychology*, 10(1), 165-172.

- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83-87.
- Furner, J. M., & Berman, B. T. (2003). Review of Research: Math Anxiety: Overcoming a Major Obstacle to the Improvement of Student Math Performance. *Childhood Education*, 79(3), 170-174.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. Review of educational research, 66(2), 99-136.
- Heit, E., & Erickson, S. (2011). In praise of secular Bayesianism. *Behavioral and Brain Sciences*, 34(04), 202-202.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (2006). Gender comparisons of mathematics attitudes and affect. *Psychology of Women Quarterly*, 14(3), 299-324
- Jakobsson, N. (2012). Gender and confidence: are women underconfident? *Applied Economics Letters*, 19(11), 1057-1059.
- Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(04), 169-188.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, 43(5), 825-832.
- Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition, *Human development*, 26, 222-232.
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219-224.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121.
- Jaccard, J., Dodge, T., & Guilamo-Ramos, V. (2005). Metacognition, risk behavior, and risk outcomes: the role of perceived intelligence and perceived knowledge. *Health Psychology*, 24(2), 161.
- 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.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18(3), 159-163.
- Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press/Bradford.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at

- predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, 2(4), 267-270.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, *36*(2), 89-101.
- Plake, B. S., & Parker, C. S. (1982). The development and validation of a revised version of the mathematics anxiety rating scale. *Educational and Psychological Measurement*, 42(2), 551-557.
- Pressley, M. (2002). Metacognition and self-regulated comprehension. What Research Has to Say About Reading Instruction, 3, 291-309.
- Pugalee, D. K. (2001). Writing, mathematics, and metacognition: Looking for connections through students' work in mathematical problem solving. *School Science and math*, 101(5), 236-245.
- Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education*, 20, 338-355.
- Schraw, G. (1996). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *The Journal of Experimental Education*, 65(2), 135-146.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*(1), 111-139.
- Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science*, *30*(4), 759-774.
- Son, L. K., & Sethi, R. (2010). Adaptive learning and the allocation of time. *Adaptive Behavior*, 18(2), 132-140.
- Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 10(2), 117-139.
- Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66.
- Townsend, C., & Heit, E. (2010). Metacognitive judgments of improvement are uncorrelated with learning rate. Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society, 1858-1862.
- Townsend, C., & Heit, E. (2011). Framing influences aggregate judgments of learning. Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society, 756-761.
- Tucker, A. B., Kelemen, C. F., & Bruce, K. B. (2001, February). Our curriculum has become math-phobic! *In ACM SIGCSE Bulletin*, *33*(1), 243-247).
- Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80(2), 210.