Journal of Applied Research in Memory and Cognition 3 (2014) 252–260

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

The effects of collaborative practice on statistical problem solving: Benefits and boundaries

Sarah Pociask, Suparna Rajaram*

Stony Brook University, Department of Psychology, Stony Brook, NY 11794-2500, United States

ARTICLE INFO

Article history:
Received 1 February 2014
Received in revised form 30 May 2014
Accepted 22 June 2014
Available online 28 June 2014

Keywords:
Collaboration
Memory
Learning
Statistics
Repeated practice

ABSTRACT

In recent years, understanding the effects of collaboration on learning and memory has emerged as a major topic of investigation. Findings from applied educational research and from basic cognitive research demonstrate a complex view of how collaboration affects learning. The present laboratory study bridged these two domains of research to address the question of how collaborative learning affects statistical problem solving. After viewing a lecture, participants completed two statistics tests. They either completed the tests collaboratively and then individually, or completed both tests individually. Results showed an immediate benefit of collaboration, but this benefit did not persist on a subsequent individual test. Repeated practice by those who worked individually increased performance to the level of those who had previously collaborated. These results were qualified by gender as females showed a consistent benefit from prior collaboration on the post-collaborative test, particularly on conceptual problems. Implications for education are discussed.

1. Introduction

Across all levels of schooling, small group instruction is one of the most popular methods of learning among both teachers and students alike; groups are organized by teachers for activities in class, as well as by students for studying and doing homework outside of class. Small group learning activities can be employed in just about every academic subject, including those that involve problem solving skills, like math and statistics (Garfield, 1993). As the prevalence of group learning methods in the classroom has increased over the past decades, the efficacy of these methods has become a major topic of research for social, applied educational, and more recently, cognitive psychologists who are interested in the role that the social context plays in shaping learning and memory performance. Scientific evidence suggests that working in groups to learn and remember information is associated with a variety of different outcomes, both positive and negative, and we are only beginning to understand the specific mechanisms that lead to these different outcomes. The present study investigates how collaborative practice affects statistical problem solving, and takes a cognitive perspective to examine it. Rather than examining verbatim memory, the present study investigates remembering in the context of problem solving, that is, how people remember what they have learned to solve particular problems. Such remembering is typical of everyday life and is one of the main contexts in which collaborative remembering often occurs.

Given the educational implications of how people learn and remember in groups, it is not surprising that a considerable amount of educational research has been conducted in this area within classrooms at all levels of schooling. With regard to statistics classes, applied educational researchers have studied the effects of implementing various cooperative learning techniques, finding a range of positive outcomes. For example, Keeler and Steinhorst (1995) found that after incorporating in-class collaborative activities into their introduction to statistics course, a higher percentage of students passed the course in comparison to past semesters when collaborative techniques were not used (also see Magel, 1998). They also found that final grade point averages improved (also see Giraud, 1997), as did student satisfaction with the course. Further, several meta-analyses and reviews have been conducted (e.g. Lei, Kuestermeyer, & Westmeyer, 2010; Lou, Abrami, & d'Apollonia, 2001; Roseth, Johnson, & Johnson, 2008; Slavin, 1980; Springer, Stanne, & Donovan, 1999) to summarize the findings from this expansive body of educational research and to determine recommendations and best-practices for teachers interested in employing collaborative techniques in their classrooms. Findings from these and other reviews and meta-analyses usually highlight the positive outcomes of collaboration such as higher levels of achievement, positive attitudes toward group-work and

---

* Corresponding author. Tel.: +1 631 632 7841.
E-mail addresses: sarah.pociask@stonybrook.edu (S. Pociask), suparna.rajaram@stonybrook.edu (S. Rajaram).

http://dx.doi.org/10.1016/j.jarmac.2014.06.005
2211-3681/

peers, positive self-esteem, better attitudes toward learning, and increased persistence in courses and programs. Consequently, collaboration is typically seen as a major success story (Johnson & Johnson, 2009) and the trend in schools is to include collaborative tasks in the classroom whenever possible.

Although these social and academic outcomes are very encouraging, it is important to note that outcomes of collaboration are not always positive (Barron, 2000, 2003; Cooper, Cox, Nammouz, Case, & Stevens, 2008; Crook & Beier, 2010; Gillies, 2003; Salomon & Globerson, 1989; Sfard & Kieran, 2001; Webb, 1982, 1993; Webb, Nemer, Chizhik, & Sugrue, 1998). Individual instances of collaboration can vary in effectiveness depending on factors related to the task (e.g. instruction, resources, complexity, goal-orientation, etc.) the learner (e.g. intelligence, learning style, social skills, etc.), and the group (e.g. group composition, size, etc.). Depending on these and other factors, results can be mixed; compared to traditional learning methods (i.e. individual learning), sometimes there are no academic or social benefits of collaboration, and other times collaboration is associated with lower performance (e.g. Crooks, Klein, Savenye, & Leader, 1998; Gadgil & Nokes-Malach, 2012; Leidner & Fuller, 1997; Slavin, 1980, for a review, see Table 3; Tudge, 1989). Furthermore, research in this discipline rarely focuses on the cognitive components of collaboration and the successes or failures for group and individual performance that might be linked to these components.

Similar to the applied education literature, laboratory research on collaborative memory has also demonstrated both positive and negative outcomes of collaboration. Research demonstrating improved memory as a result of collaboration, also known as collaborative facilitation, has found that under certain conditions groups produce more accurate and more complete accounts of their memories than if they worked independently, such as on a recognition task, working with a familiar partner (e.g., couples), experts collaborating together, or when instructed to be as accurate as possible or reach consensus as a group (e.g., Clark, Hori, Putnam, & Martin, 2000; Harris, Keil, Sutton, & Barnier, 2010; Harris, Barnier, & Sutton, 2012; Meade, Nokes, & Morrow, 2009; Yarmey & Morris, 1998). The benefits of collaboration have also been observed following group memory tasks, that is, in post-collaborative individual performance, when participants who previously remembered studied information in collaboration with group members attempt to remember the same information on their own (e.g., Blumen & Rajaram, 2008).

In contrast to these findings, group remembering is also associated with costs. For example, a classic finding in collaborative memory research is that working in groups leads to sub-optimal individual performance during group recall, a phenomenon known as collaborative inhibition (Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997). So, even though the net result for the group is greater than what any one person can remember, each individual in the group does not perform up to their full potential. This finding is consistently demonstrated across a variety of tasks, situations, and stimuli (for a review, see Rajaram & Pereira-Pasarin, 2010). Finally, the costs of collaboration can also extend to individual memory, as is seen in studies noting the social transmission of errors (e.g., Roediger, Meade, & Bergman, 2001; Thorley & Dewhurst, 2007, 2009), or post-collaborative forgetting (Basden, Basden, & Henry, 2000; Coman, Manier, & Hirst, 2009; Congleton & Rajaram, 2011).

It turns out that whether or not memory is improved by working with others depends on several interacting processes that occur during collaboration (Rajaram & Pereira-Pasarin, 2010), some that enhance probability of correct remembering (i.e. re-exposure, relearning through retrieval, and error pruning), and others that reduce that probability (i.e. social contagion of errors, blocking, and retrieval disruption). For example, serving to improve subsequent memory, hearing the responses of one's group members allows for an additional study opportunity (i.e., re-exposure). Additionally, the act of retrieving information from one's own memory can improve later recall through a process similar to rehearsal (i.e., re-learning through retrieval). Finally, by providing feedback during collaboration, group members can eliminate the production of errors in subsequent individual recall (i.e. error pruning). However, just as feedback from the group can reduce the spread of errors, group members can also make errors themselves or incorporate others' erroneous memories. If these errors go uncorrected or unchallenged, these same mistakes can persist during individual recall later, a process known as social contagion of errors (Basden et al., 1997; Roediger et al., 2001; Thorley & Dewhurst, 2007, 2009; Weldon & Bellinger, 1997). Blocking or forgetting of information may occur as a result of waiting while others contribute, or as a result of recalling what comes to mind first. After recalling all of the strongly remembered information, the weakly remembered information can no longer be accessed. Finally, since everyone has their own way of organizing and retrieving information, and hearing the information produced by someone else in a different way may disrupt one's own retrieval.

The aforementioned research largely focuses on verbal study materials (e.g., word lists, and narratives). To date, very few studies have focused on the cognitive components of collaborative learning using educationally relevant materials, and none has specifically focused on statistics learning. The statistical problem solving task in the present experiment differs from much of the traditional collaborative memory literature in that it involves both declarative or conceptual components (i.e. knowing theoretical principles and formulas) and procedural or computational components (i.e. knowing how to implement principles and formulas) (Cohen & Squire, 1980; Mestre, Ross, Brookes, Smith, & Nokes, 2009). Our goal was to investigate whether and how collaborative practice would influence statistics learning for the individual learners in this distinctly different context of learning and performance. In approaching this question, we did not intend to compare group recall and group problem solving; rather, we adapted the broad context of the collaborative memory literature and the general experimental paradigm widely used therein, to develop a controlled, laboratory test of statistics learning by individual learners. The collaborative memory literature provided a useful backdrop also because some mechanisms elaborated therein, e.g., re-exposure and relearning, are relevant for exploring possible changes in individual learning as a function of collaborative practice of statistical content in the present study.

In sum, the findings from collaborative memory research and collaborative learning paint a complex and at times conflicting view of how collaboration affects learning. Despite the considerable research on collaboration that has been conducted across various contexts, the specific effects of collaboration on content-based problem solving (e.g. statistics) remain unclear. Many of the findings within the educational literature are encouraging, but without investigating the effects of collaboration at both the group and individual level under controlled conditions, the reasons behind the demonstrated improvements will remain unspecified. That is, within an applied setting, there are often a number of practical constraints that make it difficult for some basic research design elements (e.g. proper control groups, random assignment to conditions, measurement of dependent variables, etc.) to be carried out with the same level of control that is feasible in a laboratory setting. Moreover, given what we know about the common misconceptions students hold regarding statistics topics (Garfield, 1995) and the challenges students experience with learning and applying conceptual information (Confrey, 1990; Mestre et al., 2009) more research is needed to better understand the extent to which students can benefit from collaboration in statistics courses. When considering the myriad of findings from prior research, a number of questions arise. Compared to working individually, how does collaboration affect an individual's performance? How does collaboration affect the transfer of learning across tests? Do the effects of collaboration depend on the type of test problem? Do the effects of collaboration depend on participant characteristics, such as gender?

The present study is designed within the broad scope of exploring these questions. Given that the present study aims to bridge the cognitive research on collaborative memory with the education research involving classroom subject content, its design and specific aims were shaped by considerations within each of these domains. With respect to the design, this study used an experimental protocol modeled in part after a typical collaborative memory paradigm, and also after the applied educational research conducted in classrooms. Similar to the typical collaborative memory paradigms, students in the present study first studied information alone, and were then tested on that information on two occasions; the first time students were tested either individually or in a small group, and the second time all students were tested individually (see Weldon & Bellinger, 1997). This comparison creates a systematic and direct experimental test that is often lacking in classroom settings. In particular, it enables a test of whether test practice repeated individually can improve performance and, if it does, whether it can elevate an individual's problem solving performance to the same level as would prior collaborative practice. Following from applied educational research, the present study involves educationally relevant materials (e.g. a statistics lecture and problem sets) and tasks designed to be similar to those that would be experienced in a college course. Importantly, we examine problem solving because it is a more ecologically valid measure of what people remember from the lecture that a rote or verbatim memory test. In addition to following the sequence of a collaborative memory paradigm, the present design reflects a common sequence of learning as it takes place in the classroom; first the student absorbs the information alone through listening to a lecture, then the student may work individually or in a small group to study (e.g. through an in-class assignment or homework), and finally the student is tested on the material individually on a quiz or an exam. By merging these two approaches, we hope to be able to further understand the effects of collaborative practice on individual learning, specifically with respect to statistical problem solving.

2. Method

2.1. Participants

A total of 192 first and second year undergraduates—an equal number of males and females—at Stony Brook University participated. Each experimental session consisted of only males or only females in order to minimize the influence of stereotype threat (Huguet & Régner, 2007).

2.2. Design

There were three independent variables, which included experimental condition (individual or collaborative), problem type (near transfer, far transfer, and conceptual; explained below and in Fig. 1), and gender. The primary dependent variable was performance on the two consecutive problem solving tests, which was assessed by calculating the proportion of correct responses for each participant on each of the two tests. Other dependent variables included prior educational background, source of knowledge ratings, ratings of the experimental tasks, and preferences for collaboration (all described in the following subsections). A measure of anxiety was taken at several points during the experiment, i.e., the State-Trait Anxiety Inventory, (STAI; Spielberger, 1983); as these measures did not differ in any of the comparisons at any time point, for brevity these findings are not reported.

2.3. Materials

2.3.1. Lecture

During the study phase, participants viewed a 30-min lecture about the computational and conceptual aspects of the statistical concept of central tendency. It included information about how to compute the mean, median, and mode, as well as information about why multiple measures of central tendency exist and when it is recommended to use each measure. The lecture was a pre-recorded PowerPoint presentation with a synchronized audio voiceover (female voice) explaining the content on the slides. Additional details regarding the lecture are available upon request from the corresponding author.

2.3.2. Statistics tests

Participants completed two consecutive tests on the material presented in the lecture. Each test contained three types of problems: near transfer, far transfer, and conceptual problems (see Fig. 1). Near and far transfer problems both involve computations. Near transfer problems are presented in the exact same format as they were in the lecture, whereas far transfer problems are displayed in a different format or involve a different application of a formula (e.g., Paas, 1992). Conceptual questions do not necessarily involve any computations, but rather pertain to broad, overarching ideas and may involve comparing multiple concepts. Each of the two tests consisted of its own unique set of problems; however, all problems were matched across tests so the same topics appeared in both versions. Both sets of problems were counterbalanced to appear in each test.

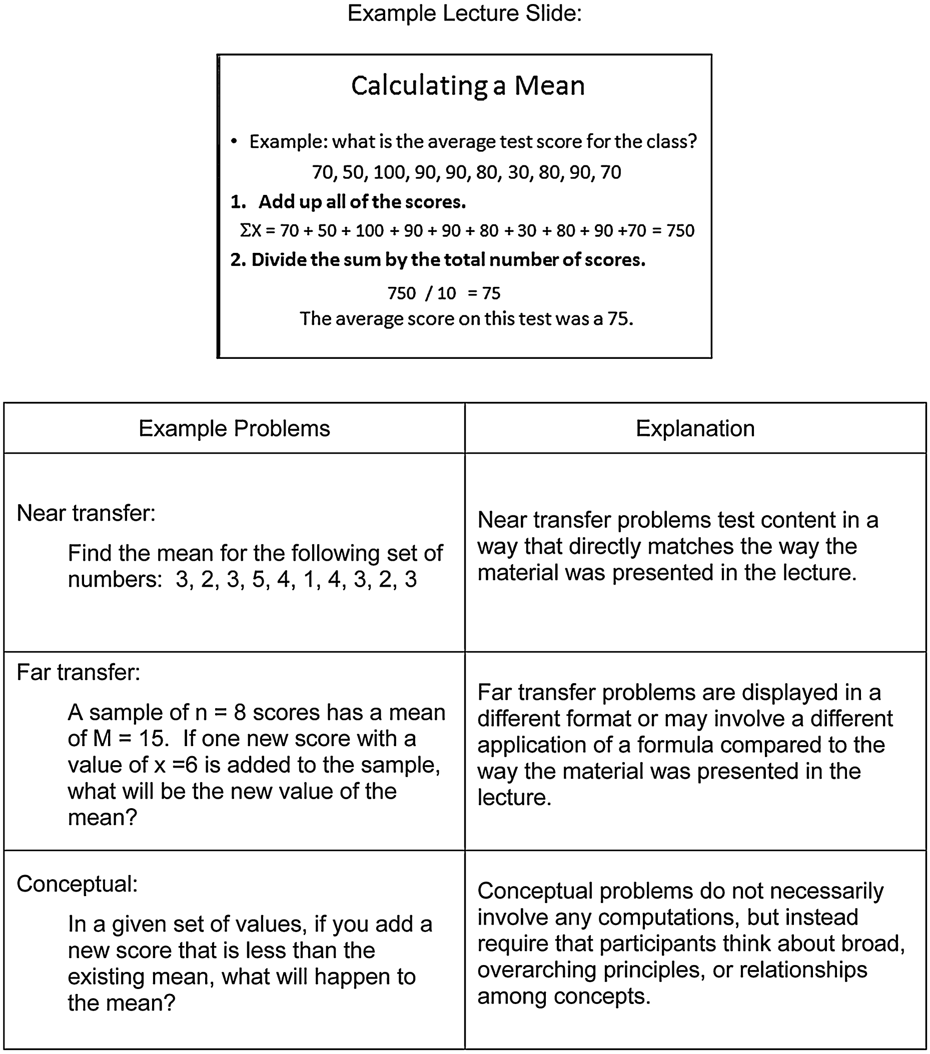
The problems were presented one at a time, both on the computer screen and also on paper in the test packet, and participants had one and a half minutes to complete each problem. When it was time to move onto the next problem, participants heard a tone, and the next problem appeared on the screen. After hearing the tone, participants were told to write down a final answer in the test packet if they had one, and then immediately move onto the next question. If participants finished a problem early, they were told to press the space bar to advance to the next problem. During the test, problems were completed in a forward sequence only. Extensive pilot testing was carried out to develop and calibrate the type and selection of problems, the matched problem sets that appeared across the two tests, and the procedure used for subjects to solve the problems.

2.3.3. Source of knowledge ratings

All participants completed source of knowledge ratings after finishing the first test. For each problem completed, participants were asked to indicate the source of their knowledge, or in other words, how they knew how to solve the problem. Additional details regarding these ratings can be found in Section 3.

2.3.4. Task impressions and collaboration preferences questionnaire

In order to understand how experimental tasks were perceived, all participants were asked to provide feedback on the study phase and test phases at the end of the experimental session. Additionally, all participants answered questions about their beliefs and preferences for collaboration. As these findings did not shape or alter the patterns of results reported in this manuscript, for the sake of economy these questions and findings are not discussed further.



Example Lecture Slide:

Calculating a Mean

- Example: what is the average test score for the class?

70, 50, 100, 90, 90, 80, 30, 80, 90, 70

1. Add up all of the scores.

$\Sigma X = 70 + 50 + 100 + 90 + 90 + 80 + 30 + 80 + 90 + 70 = 750$

2. Divide the sum by the total number of scores.

$750 / 10 = 75$

The average score on this test was a 75.

Example Problems	Explanation
Near transfer: Find the mean for the following set of numbers: 3, 2, 3, 5, 4, 1, 4, 3, 2, 3	Near transfer problems test content in a way that directly matches the way the material was presented in the lecture.
Far transfer: A sample of n = 8 scores has a mean of M = 15. If one new score with a value of x =6 is added to the sample, what will be the new value of the mean?	Far transfer problems are displayed in a different format or may involve a different application of a formula compared to the way the material was presented in the lecture.
Conceptual: In a given set of values, if you add a new score that is less than the existing mean, what will happen to the mean?	Conceptual problems do not necessarily involve any computations, but instead require that participants think about broad, overarching principles, or relationships among concepts.

Fig. 1. Example of lecture content and near transfer, far transfer, and conceptual problems.

2.3.5. *Educational background questionnaire*

In order to assess participants' educational background and proficiency in math and statistics, especially given the between-subjects assignment of subjects, at the conclusion of the study participants were asked to provide their high school GPA, SAT scores, and their prior coursework in math and statistics in both high school and college.

2.4. *Procedure*

The experiment consisted of the following phases in sequence: friendship screening, initial completion of the STAI-S, study phase (i.e. lecture viewing), delay, first test phase, second completion of the STAI-S, source of knowledge ratings, delay, second test phase, the third completion of the STAI-S, and task impressions, collaboration preferences, and educational background questionnaires.

At the beginning of the experimental session, participants who knew each other in some way were identified to construct groups consisting of only strangers. Participants who did not know each other were randomly assigned to either complete the first test individually, or in a collaborative triad, where they could discuss the problems as a group before writing down their final individual answers.

Next, participants completed the STAI-S, and then each participant was seated at a computer and was told that s/he would be viewing a short lecture that contained material s/he would be tested on in subsequent phases of the experiment.

Following the lecture, participants were given a 15-min break during which they played a game on the computer. Immediately after the break, participants received instructions for the first test phase. If completing the first test individually, participants were assigned to sit at their own computers and complete the problems without interacting with other participants. If completing the first test collaboratively, all three participants sat at one computer, and were told that they would be working with their partners to solve the problems in the test. Even though they were working as a group, each participant received his/her own test packet, and was responsible for writing down his/her own answers (i.e. they did not need to reach consensus). To simulate the typical group study procedure in classrooms, groups were allowed to discuss their own strategies for solving each problem and were not given any specific goals or instructions for how to work together. The only requirement for the collaboration was that all group members work on the same problem at the same time.

All participants were told that they would be tested twice. Those completing the first test individually were told that they would complete both tests on their own, and those in the collaborative group were told that they would complete the first test together, and the second test individually. A basic calculator was also provided for use during both tests.

Following the first test, participants completed the STAI-S once again, followed by the source of knowledge ratings, and were then given another 15-min break before the second test phase. At the start of the second test phase, all participants were seated at their own individual computers and were informed that the second test contained a new set of questions, but that all other aspects of the test were exactly the same as the first (e.g. timing, format, etc.).

After the second test phase, participants completed the end of experiment questionnaires (one final administration of the STAI-S, a series of questions about the experimental tasks and preferences for collaboration, and an educational background questionnaire). Participants were then debriefed and asked not to discuss the experiment with others. The entire experimental session took about 3 h to complete.

3. Results

3.1. Test performance

The primary goals of the present study were to assess the immediate and post-collaborative effects of collaboration on performance on consecutive statistical problem solving tests, and moreover, to investigate whether these effects are consistent across problem type (i.e. near transfer, far transfer, and conceptual problems). Because we controlled the group composition for gender by implementing an equal number of all-male and all-female triads, we also examined performance as a function of gender. In all conditions, the test performance was assessed at the individual, and not the group, level.

For the analyses, pairwise comparisons were Bonferroni-corrected to the .05 level, and the Greenhouse–Geisser correction was used for any violations of sphericity in our analyses with repeated measures factors.

3.1.1. Immediate and post-collaborative effects

First, in order to assess the immediate effects of collaboration, we conducted a one-way analysis of variance (ANOVA) on the proportion of correct responses on the first test with condition (collaborative, individual) as a between subjects factor. There was an immediate benefit of collaboration, $F(1,190) = 13.92$, $p < .001$, $\eta_p^2 = .07$, such that those who worked collaboratively on the first test performed significantly better ($M = .75$, $SE = .01$) than those who worked individually ($M = .69$, $SE = .01$). However, looking at performance on the second test where all participants worked individually, we see that this advantage for collaboration did not persist, $F(1,190) = .10$, $p = .76$, $\eta_p^2 = .001$. Individual performance on the second test was the same regardless of whether participants had previously collaborated or worked individually ($M = .74$, $SE = .01$, in both conditions).

Comparing across the consecutive tests, individual performance was consistent for those who first worked collaboratively, $t(95) = .96$, $p = .34$, $d = -.11$, indicating that participants maintained the initial benefit from collaboration when they completed the second test individually. However, participants who completed both tests individually showed a significant improvement from the first to second test, $t(95) = -6.40$, $p < .001$, $d = .38$. This improvement in performance is similar to hypermnesia effects in verbal recall (Payne, 1987) where recall levels improve with repeated attempts at recalling once-studied information. Here, the problems and questions themselves differed across the two tests but probed the same knowledge base that was provided at study. Thus, our first key finding revealed that repeated retrieval or, in the present case, repeated attempts at solving statistics problems by oneself, enhanced performance to the same level as prior collaboration.

Table 1
Mean proportion of correct responses as a function of problem type and condition (standard error in parentheses).

	Collaborative *M* (SE)	Individual *M* (SE)
First test		
Near transfer	0.97 (.005)	0.93 (.01)
Far transfer	0.72 (.02)	0.66 (.02)
Conceptual	0.63 (.01)	0.57 (.02)
Second test		
Near transfer	0.95 (.007)	0.93 (.01)
Far transfer	0.72 (.02)	0.74 (.02)
Conceptual	0.62 (.01)	0.60 (.02)

3.1.2. Problem type

A second question of interest was whether these effects depend on the type of problem. After calculating the proportion of correct responses for near transfer, far transfer, and conceptual problems, we conducted a 2 × 3 mixed ANOVA with condition (collaborative, individual) and problem type (near transfer, far transfer, and conceptual) for each test. As expected, on both tests there was a significant main effect for problem type such that performance was the best on near transfer problems, followed by far transfer problems, and performance was the worst on the conceptual problems (all p values less than .001), see Table 1. Demonstrating a consistent benefit of collaboration for all three problem types on the first test, there was a significant main effect of condition, $F(1,190) = 15.78$, $p < .001$, $\eta_p^2 = .08$, and critically, no condition × problem type interaction, $F(2,380) = .21$, $p = .80$, $\eta_p^2 = .001$. On the second test, the collaborative advantage over individual performance did not persist for any of the three problem types, as there was no main effect for condition, $F(1,190) = .31$, $p = .58$, $\eta_p^2 = .002$, and no condition × problem type interaction, $F(2,380) = 1.42$, $p = .24$, $\eta_p^2 = .007$.

3.1.3. Gender

In order to see if the effects of collaboration depended on gender, we conducted a 2 × 2 ANOVA of gender (male, female) and condition (collaborative, individual) on the proportion of correct responses on both tests. Given that the Levene's Test was significant for both ANOVAs, we used the recommendation of a more conservative alpha level of .025 for these analyses.

On the first test, the main effect of condition was significant, $F(1,188) = 14.12$, $p < .001$, $\eta_p^2 = .07$, and there was no gender × condition interaction, $F(1,188) = .37$, $p = .55$, $\eta_p^2 = .002$, demonstrating that the benefit of collaboration on the first test was observed consistently for both males and females (see Table 2). Although males performed numerically higher than females, the main effect of gender was only marginally significant, $F(1,188) = 4.25$, $p = .04$, $\eta_p^2 = .02$.

Looking at post-collaborative effects on the second test, the main effect of condition was not significant, $F(1,188) = .10$, $p = .75$, $\eta_p^2 = .001$, nor was the main effect for gender, $F(1,188) = 3.63$, $p = .06$, $\eta_p^2 = .02$, however, there was a significant interaction between gender and condition, $F(1,188) = 6.19$, $p = .01$, $\eta_p^2 = .03$ (see Table 2). What appears to be driving the interaction is the difference in performance between males and females who had previously worked individually compared to the difference between males and

Table 2
Mean proportion of correct responses for male and female participants working collaboratively or individually (standard error in parentheses).

	Collaborative		Individual	
	Male *M* (SE)	Female *M* (SE)	Male *M* (SE)	Female *M* (SE)
First test	.76 (.01)	.74 (.01)	.72 (.02)	.67 (.02)
Second test	.74 (.01)	.75 (.01)	.77 (.02)	.70 (.02)

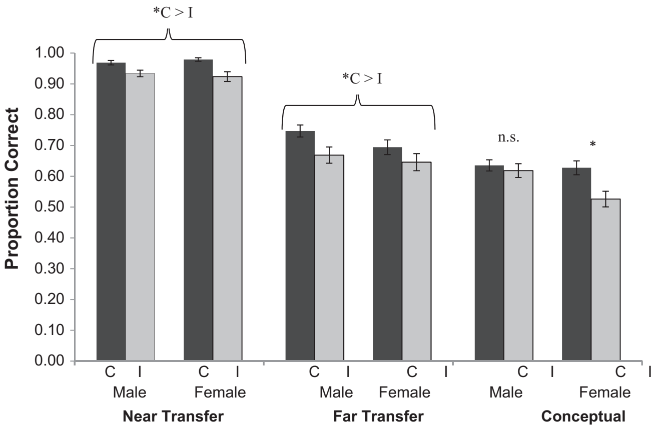


Fig. 2. First test performance as a function of gender, condition, and problem type. Males and females consistently benefit from collaboration on near and far transfer problems, but on conceptual problems, this benefit is only observed for females.

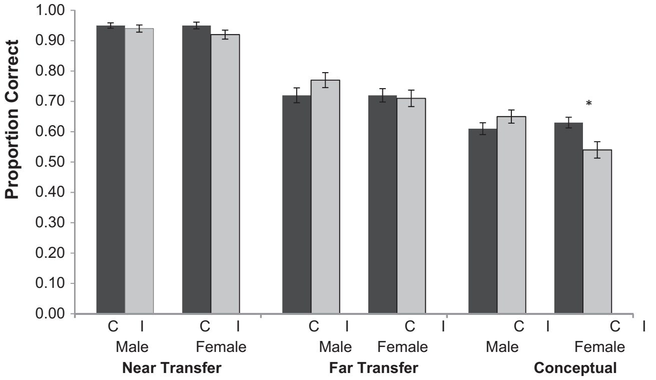


Fig. 3. Second test performance as a function of gender, condition, and problem type. On the second test there is no collaborative benefit for near and far transfer problems, but on conceptual problems, there is a persistent benefit of collaboration that is only observed for females.

females when they had previously collaborated.[1] If they had collaborated on the first test, males and females performed at the same level on the second individual test, $t(94) = -.48$, $p = .63$, $d = -.11$, but if they had worked individually on the first test, males performed significantly better than females on the second test, $t(94) = 2.76$, $p = .007$, $d = .58$. In other words, the gender difference between males and females on the second test is not observed for participants who had previously worked collaboratively.

3.1.4. Gender and problem type

Finally, in order to further understand the nature of the gender × condition interaction, we conducted a 2 (gender) × 2 (condition) × 3 (problem type) mixed ANOVA on the proportion of correct responses on both tests to see if the effects of gender and collaboration depended on problem type. Looking at performance on the first test, there was a significant main effect of problem type, $F(2,376) = 442.80$, $p < .001$, $\eta_p^2 = .70$ and a significant main effect of condition, $F(1,188) = 16.00$, $p < .001$, $\eta_p^2 = .08$, but most critically, there was a marginal gender × condition × problem type interaction, $F(2,376) = 2.84$, $p = .06$, $\eta_p^2 = .015$. A series of gender × condition interaction contrasts revealed a significant main effect of condition for both near and far transfer problems ($F(1, 188) = 20.91$, $p < .001$, $\eta_p^2 = .10$; $F(1, 188) = 6.78$, $p = .01$, $\eta_p^2 = .04$, respectively), and no gender × condition interaction (both Fs < 1). For conceptual problems, there was also a significant main effect of condition, $F(1, 188) = 7.23$, $p < .01$, $\eta_p^2 = .04$, however, this was qualified by a marginal gender × condition interaction, $F(1, 188) = 3.51$, $p = .06$, $\eta_p^2 = .02$. Follow-up comparisons reveal that while there was no collaborative benefit for males on conceptual problems, $t(94) = .63$, $p = .53$, $d = .14$, females did benefit, $t(94) = 3.00$, $p < .01$, $d = .59$. In summary, these analyses show that while both males and females consistently benefited from collaboration on near and far transfer problems on the first test, only females benefited on conceptual problems (see Fig. 2).

On the second test, the gender × condition × problem type interaction was significant, $F(2,376) = 3.12$, $p < .05$, $\eta_p^2 = .016$. Again we see that the pattern of performance is consistent for males and females on near and far transfer problems, where there are no differences in performance for males and females who had previously collaborated or worked individually (all p-values greater than .13). However, looking at performance on conceptual problems, females show a persistent benefit of collaboration on the second test, $t(94) = -2.91$, $p < .01$, $d = .57$, whereas males do not, $t(94) = 1.64$, $p = .11$, $d = -.27$ (see Fig. 3).

3.2. Educational background

One possible explanation for the observed differences in male and female performance is that there were differences in prior educational experiences and academic achievement across condition and gender that could account for differences in test performance. However, this was not the case. Self-reported high school GPA, SAT scores, total number of high school math classes, and total number of college math classes were analyzed using a series of 2 × 2 ANOVAs with condition (collaborative, individual) and gender as between subjects factors. The few significant main effects that were observed favored the underperforming groups such that females ($M = 3.63$, $SE = .03$), reported higher high school GPAs than males ($M = 3.45$, $SE = .04$), $F(1, 180) = 12.57$, $p < .001$ $\eta_p^2 = .07$; and participants in the individual condition ($M = 1.46$, $SE = .09$), reported taking more college math classes than those in the collaborative condition ($M = 1.17$, $SE = .08$), $F(1, 188) = 5.45$, $p = .02$ $\eta_p^2 = .03$ (see Table 3 for a summary of these results). Furthermore, none of the condition × gender interactions were significant for any of these educational background questions. It is additionally worth noting that the majority of participants (75%) had not taken a stats class in the past 6 months, and this distribution was the same across all conditions, $\chi^2(3, N = 191) = 1.29$, $p = .73$. Therefore, it is unlikely that differences in prior educational experiences and academic achievement account for the observed pattern of test performance results in the present study.

3.3. Source of knowledge

After completing the first test, all participants indicated the source of knowledge for each question that they answered. Participants selected whether they felt that they knew how to solve the problem before the experiment, whether they learned it from the lecture, or if they simply guessed; and for participants who worked collaboratively, whether they learned how to solve it from working in their group. It is important to note that the source of knowledge ratings do not reflect an objective assessment of prior knowledge, and were not intended for this purpose. Rather, these ratings reflect perceived prior knowledge and index the participants' own self-assessment of how they knew how to solve the problem. Participants were allowed to select multiple sources of

---

[1] Numerically, males who had previously worked collaboratively performed slightly worse than males who had previously worked individually and females who first worked collaboratively performed slightly better than females who first worked individually (Table 3), but these differences did not reach statistical significance ($t(94) = 1.68$, $p = .10$, $d = .30$; $t(94) = -1.84$, $p = .07$, $d = -.42$, respectively).

Table 3
Educational background as a function of gender and collaborative condition (means with standard error in parentheses).

	Collaborative		Individual		Significance
	Male *M* (*SE*)	Female *M* (*SE*)	Male *M* (*SE*)	Female *M* (*SE*)	
High school GPA	3.46 (.06)	3.61 (.05)	3.44 (.05)	3.64 (.04)	*p* = .001; female > male
Composite SAT (verbal and quantitative)	1232.73 (19.12)	1238.50 (23.63)	1225.48 (20.60)	1200.25 (24.49)	n.s.
Total high school math classes	5.87 (.29)	5.71 (.28)	6.15 (.26)	5.77 (.30)	n.s.
Total college math classes	1.23 (.12)	1.10 (.11)	1.48 (.13)	1.44 (.14)	*p* = .02; individual > collaborative

knowledge for each problem; however these instances were rare (less than 1% of all problems). For clarity of interpretation, the following analyses include only problems where only one source of knowledge was selected (the results remain the same in both analyses). A summary of the proportion of problems attributed to each source as a function of problem type, condition, and gender can be found in Table 4.

First, we will focus on the proportion of problems where participants indicated that they already knew how to solve the problem before the experiment. Overall, males said that they already knew how to solve the problem before the experiment more frequently than females across all problem types, $F(1,188) = 27.95$, $p < .001$, $\eta_p^2 = .13$, and given the similar educational backgrounds of the male and female participants (see Table 3), this pattern may reflect greater confidence in their pre-existing knowledge.

Next, looking at the proportion of problems where participants indicated that they learned how to solve the problem from the lecture, females said that they learned how to solve the problem from the lecture more frequently than males across all problem types, $F(1,188) = 13.26$, $p < .001$, $\eta_p^2 = .07$. In addition to main effect of gender, there was also a main effect of condition such that participants who worked individually said that they learned how to solve the problem from the lecture more frequently compared to those who had worked collaboratively, $F(1,188) = 26.18$, $p < .001$, $\eta_p^2 = .12$.

Similarly, there was a main effect of gender for guessing, $F(1,188) = 19.51$, $p < .001$, $\eta_p^2 = .09$, such that females said that they guessed on more problems than males, and a main effect of condition, $F(1,188) = 68.12$, $p < .001$, $\eta_p^2 = .27$, such that those working individually guessed more than those working collaboratively. However, it is important to note that this pattern of results is qualified by a gender × condition interaction, $F(1,188) = 22.05$, $p < .001$, $\eta_p^2 = .11$, where males and females reported similar levels of guessing if they had collaborated, $t(94) = .30$, $p = .76$, $d = .06$; but if they worked individually, females reported a significantly higher level of guessing than males, $t(94) = -5.13$, $p < .001$, $d = 1.12$.

Finally, when looking at the proportion of problems attributed to the group as the source of knowledge, males and females who worked collaboratively attributed equal number of problems to the group for near and far transfer problems ($t(94) = 1.18$, $p = .24$, $d = .23$.; and $t(94) = .80$, $p = .43$, $d = .15$, respectively), which parallels the test performance data where males and females both showed a collaborative benefit on these types of problems. Interestingly, for conceptual problems females who worked collaboratively reported that they learned how to solve the problem by working as a group more frequently than did males, $t(94) = 2.40$, $p < .05$, $d = .49$, which again parallels test performance data where females benefited from collaboration on conceptual problems, but males did not.

4. Discussion

The results demonstrate a clear immediate advantage of collaborating on the students' ability to solve statistics problems. Participants—both males and females—who collaborated before individually answering questions on the first test outperformed those who worked only individually, and this level of performance was maintained on the second test when everyone worked on alone. However, despite performing at a lower level on the first test, participants who worked only individually increased their performance on the second test, equal to the level of their peers who had first worked collaboratively. Thus, due to the increased performance for those who worked individually, the initial advantage for collaboration that was observed after the first test did not persist on the second test. In other words, the downstream benefits of collaboration did not exceed those that accrued from repeated test practice for participants who worked alone throughout. We return to this broad finding later.

Interestingly, these effects depended to some extent on gender and problem type. The initial benefit of collaboration was observed for both males and females on problems that tested computational aspects of statistical problem solving, or knowing how, (i.e. near transfer and far transfer problems). However, for problems that tested the conceptual aspects of statistics, or knowing that, females benefited from collaboration, but males did not. Moreover, although there was no post-collaborative benefit observed for males or females on near and far transfer problems on the second test, females showed a persistent benefit of collaboration on conceptual problems. Finally, it is also critical to note that the differences in performance related to prior collaboration and gender were not explained by differing educational backgrounds or anxiety.

So, why is it that females consistently benefit from collaboration, whereas males do not show this benefit in all instances? The answer to this question cannot be fully answered with results from the present study, but some patterns in the present findings may present conjectures for future investigations. For instance, even though males and females had similar educational backgrounds, the source of knowledge data showed that, compared to females, males were more likely to report that they knew before the experiment how to solve the problems. This belief might have reduced the reliance on collaboration for males compared to females who, by comparison, reported more guessing. Since conceptual problems require that participants explain and relate concepts to one another rather than directly apply formulas (Garfield & Ben-Zvi, 2007; Mestre et al., 2009), reliance on collaboration may be especially important for these types of problems. However, this possibility about the observed relationship in our findings, or other possible explanations, require follow-up in future work.

Regardless of why collaboration additionally helped females in our study, the present findings suggest that collaborative learning may be an important educational tool to help female students succeed in statistics and other STEM courses. This is particularly relevant given the underrepresentation of women in STEM domains (St. Rose, 2010; Syed & Chemers, 2011), the risk of underperformance in math-related domains due to stereotype threat, (Elizaga & Markman, 2008; Nguyen & Ryan, 2008; Rydell, Rydell, & Boucher, 2010; Spencer, Steele, & Quinn, 1999), and the report that females prefer to engage in activities involving communal goals even when their natural abilities are equally strong in STEM domains (Diekman, Brown, Johnston, & Clark, 2010). As such, further research examining the long-term effects of collaborative activities in STEM domains is warranted.

Table 4
Mean proportion of problems attributed to each source as a function of problem type, condition, and gender (standard error in parentheses).

	Near transfer				Far transfer				Conceptual			
	Male		Female		Male		Female		Male		Female	
	C[a]	I	C	I	C	I	C	I	C	I	C	I
Knew	0.93 (.01)	0.93 (.01)	0.88 (.03)	0.83 (.03)	0.73 (.03)	0.86 (.02)	0.68 (.03)	0.72 (.03)	0.66 (.03)	0.68 (.03)	0.54 (.04)	0.38 (.04)
Learned	0.04 (.01)	0.05 (.01)	0.08 (.02)	0.14 (.02)	0.01 (.003)	0.05 (.01)	0.01 (.005)	0.11 (.02)	0.12 (.02)	0.20 (.03)	0.16 (.02)	0.31 (.03)
Guess	0.01 (.004)	0.02 (.01)	0.01 (.004)	0.03 (.01)	0.05 (.01)	0.09 (.01)	0.06 (.01)	0.18 (.02)	0.06 (.01)	0.13 (.02)	0.04 (.01)	0.32 (.03)
Group	0.02 (.01)	–	0.04 (.02)	–	0.22 (.03)	–	0.25 (.03)	–	0.17 (.03)	–	0.26 (.03)	–

[a] C, collaborative; I, individual.

Returning to the first main finding we discussed, the improved performance from the first test to the second observed for participants who completed both tests individually presents an important anchor against which to assess whether collaboration benefits downstream individual performance over and above individual test practice. Past research on collaborative memory has explored this question for verbal materials, which constitutes declarative memory (or knowing that). This body of research has repeatedly shown a post-collaborative advantage, or higher recall following collaboration that exceeds hypermnesia effects in repeated individual recall. In our study, we explored the effects of collaboration on study content that combines declarative and procedural (knowing how) components of learning and memory. For this learning, post-collaborative benefits did not uniformly exceed the benefits that accrue from equivalent efforts at individual practice. Because the post-collaborative benefits were not widespread for all participants and all types of problems in the present study, future work may be directed to exploring the specific variables and conditions that optimize such benefits in statistical problem solving.

Finally, the time course of the observed collaborative advantage in the present study provides one possible explanation for why students and educators so often advocate for group work. Because there are immediate advantages in performance from collaboration, this experience of efficacy likely reinforces the social benefits of working with others while performing challenging tasks. Furthermore, our findings also revealed sustained benefits in post-collaborative performance such that performance did not drop below the level of prior collaboration session even when participants worked individually on a different set of problems. This sustained level can further reinforce the experience of benefit from collaborative practice. Although individuals working alone can quickly reach the same level of performance in statistical problem solving, the engagement process in a group setting likely leads to the impression that collaboration actually confers greater advantage. To the extent this impression can motivate students to engage in more learning and practice compared to what they would do if working alone, collaboration would still be a useful pedagogical tool.

In sum, our findings show that the nature of subject matter, the type of test probes, and the type of student all contribute to determine the effects of collaboration on learning. Systematic consideration of these and other key factors related to learning and instruction can provide the bases for tailoring pedagogical tools.

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, *9*(4), 403–436.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359.

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L., III. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1176–1189. http://dx.doi.org/10.1037/0278-7393.23.5.1176

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, *14*(6), 497–507.

Blumen, H. M., & Rajaram, S. (2008). Influence of re-exposure and retrieval disruption during group collaboration on later individual recall. *Memory*, *16*(3), 231–244.

Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group collaboration in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1578–1588. http://dx.doi.org/10.1037/0278-7393.26.6.1578

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*, 207–210.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation socially shared retrieval-induced forgetting of september 11 memories. *Psychological Science*, *20*(5), 627–633.

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. *Review of Research in Education*, *16*, 3–56.

Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory,. *Journal of Experimental Psychology: General*, *140*(4), 535.

Cooper, M. M., Cox, C. T., Jr., Nammouz, M., Case, E., & Stevens, R. (2008). An assessment of the effect of collaborative groups on students' problem-solving strategies and abilities. *Journal of Chemical Education*, *85*(6), 866.

Crook, A. E., & Beier, M. E. (2010). When training with a partner is inferior to training alone: The importance of dyad type and interaction quality. *Journal of Experimental Psychology: Applied*, *16*(4), 335–348. http://dx.doi.org/10.1037/a0021913

Crooks, S. M., Klein, J. D., Savenye, W., & Leader, L. (1998). Effects of cooperative and individual learning during learner-controlled computer-based instruction. *Journal of Experimental Education*, *66*(3), 223–244. http://dx.doi.org/10.2307/20152560

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles a new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, *21*(8), 1051–1057.

Elizaga, R., & Markman, K. (2008). Peers and performance: How in-group and out-group comparisons moderate stereotype threat effects. *Current Psychology*, *27*(4), 290–300. http://dx.doi.org/10.1007/s12144-008-9041-y

Gadgil, S., & Nokes-Malach, T. J. (2012). Overcoming collaborative inhibition through error correction: A classroom experiment. *Applied Cognitive Psychology*, *26*(3), 410–420. http://dx.doi.org/10.1002/acp.1843

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*(1).

Garfield, J. (1995). How students learn statistics. *International Statistical Review/Revue Internationale de Statistique*, *63*(1), 25–34. http://dx.doi.org/10.2307/1403775

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–396.

Gillies, R. M. (2003). The behaviors, interactions, and perceptions of junior high school students during small-group learning. *Journal of Educational Psychology*, *95*(1), 137–147. http://dx.doi.org/10.1037/0022-0663.95.1.137

Giraud, G. (1997). Cooperative Learning and Statistics Instruction. *Journal of Statistics Education*, *5*.(3).

Harris, C., Keil, P., Sutton, J., & Barnier, A. (2010). Collaborative remembering: When can remembering with others be beneficial? In *Paper presented at the 9th conference of the Australasian Society for Cognitive Science*. Sydney: Macquarie Centre for Cognitive Science.

Harris, C. B., Barnier, A. J., & Sutton, J. (2012). Consensus collaboration enhances group and individual recall accuracy. *Quarterly Journal of Experimental Psychology*, *65*(1), 179–194.

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology*, *99*(3), 545–560. http://dx.doi.org/10.1037/0022-0663.99.3.545

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379. http://dx.doi.org/10.3102/0013189x09339057

Keeler, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, *3*(2).

Lei, S. A., Kuestermeyer, B. N., & Westmeyer, K. A. (2010). Group composition affecting student interaction and achievement: Instructors' perspectives. *Journal of Instructional Psychology*, *37*(4), 317–325.

Leidner, D. E., & Fuller, M. (1997). Improving student learning of conceptual information: GSS supported collaborative learning vs. individual constructive learning. *Decision Support Systems*, *20*(2), 149–163.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*(3), 449–521. http://dx.doi.org/10.3102/00346543071003449

Magel, R. C. (1998). Using Cooperative Learning in a Large Introductory Statistics Class. *Journal of Statistics Education*, *6*.(3).

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory*, *17*(1), 39–48.

Mestre, J. P., Ross, B. H., Brookes, D. T., Smith, A. D., & Nokes, T. J. (2009). How cognitive science can promote conceptual understanding in physics classrooms. In I. M. Saleh, & M. S. Khine (Eds.), *Fostering scientific habits of mind: Pedagogical knowledge and best practices in science education* (pp. 145–171).

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. http://dx.doi.org/10.1037/a0012702

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*(1), 5–27. http://dx.doi.org/10.1037/0033-2909.101.1.5

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, *5*(6), 649–663. http://dx.doi.org/10.1177/1745691610388763

Roediger, H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*(2), 365–371.

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin*, *134*(2), 223–246. http://dx.doi.org/10.1037/0033-2909.134.2.223

Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology*, *99*(6), 883–896. http://dx.doi.org/10.1037/a0021139

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, *13*(1), 89–99. http://dx.doi.org/10.1016/0883-0355(89)90018-9

Sfard, A., & Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students' mathematical interactions. *Mind, Culture & Activity*, *8*(1), 42–76.

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, *50*(2), 315–342.

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. http://dx.doi.org/10.1006/jesp.1998.1373

Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory (form Y)*. Palo Alto, CA: Mind Garden.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51.

Rose, St. A. (2010). STEM major choice and the gender pay gap. *On Campus With Women*, *39*(1), 1.

Syed, M., & Chemers, M. M. (2011). Ethnic minorities and women in STEM: Casting a wide net to address a persistent social problem. *Journal of Social Issues*, *67*(3), 435–441. http://dx.doi.org/10.1111/j.1540-4560.2011.01708.x

Thorley, C., & Dewhurst, S. A. (2007). Collaborative false recall in the DRM procedure: Effects of group size and group pressure. *European Journal of Cognitive Psychology*, *19*(6), 867–881.

Thorley, C., & Dewhurst, S. A. (2009). False and veridical collaborative recognition. *Memory*, *17*(1), 17–25.

Tudge, J. (1989). When collaboration leads to regression: Some negative consequences of socio-cognitive conflict. *European Journal of Social Psychology*, *19*(2), 123–138.

Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology*, *74*(5), 642–655. http://dx.doi.org/10.1037/0022-0663.74.5.642

Webb, N. M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment*, *1*(2), 131.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal*, *35*(4), 607–651.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1160–1175. http://dx.doi.org/10.1037/0278-7393.23.5.1160

Yarmey, A. D., & Morris, S. (1998). The effects of discussion on eyewitness memory. *Journal of Applied Social Psychology*, *28*(17), 1637–1648. http://dx.doi.org/10.1111/j.1559-1816.1998.tb01693.x