

An Induced Successful Performance Enhances Student Self-Efficacy and Boosts Academic Achievement

Akitoshi Uchida

Sairyō Junior High School

Robert B. Michael 

University of Louisiana at Lafayette

Kazuo Mori 

Tokyo University of Agriculture and Technology
Matsumoto University

A growing body of correlational research finds a relationship between self-efficacy—beliefs in one’s capabilities—and academic success. But few studies have investigated whether self-efficacy is causally related to academic success. We hypothesized that an experience of success would promote self-efficacy in junior high school students and would lead to academic improvements. To induce an experience of success, we secretly presented easy anagrams to target students (41 males and 43 females; 12–13 years old) who then outperformed their classmates (116 males and 115 females). We assessed students’ self-efficacy and academic achievement scores before and after the anagram tasks. We found that the success-induced students raised their self-efficacy, and this elevated self-efficacy persisted for as long as one year. Moreover, success-induced males eventually showed significant improvement in their academic achievement. These results provide a real-world experimental enactment of Bandura’s self-efficacy theory and have implications for the practices of educational practitioners.

Keywords: *academic achievement, an RCT experiment, motivation, schools, sex differences*

Introduction

Believing in our ability to succeed matters. Students who more strongly endorse these beliefs of *self-efficacy* are better able to monitor their activities, adopt proximal goals, select well-tuned strategies, and motivate themselves (Bandura, 1977, 1986, 1997). Bandura (1997) defined self-efficacy as “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (p. 3). It is little wonder, then, that many educational psychologists have investigated the role of self-efficacy in learning (see Pajares & Schunk, 2001; and van Dinther, Dochy, & Segers, 2011, for a review). One overarching question drives much of this research: To what extent does the strengthening of self-efficacy improve academic achievement?

At first glance, it might seem as though psychological research has successfully answered this question. Several

studies, for example, have demonstrated positive relationships among self-efficacy, academic achievement, and other intervening variables (e.g., Britner & Pajares, 2006; Kupermintz, 2002; Lau & Roeser, 2002; Pajares, Britner, & Valiante, 2000; Usher & Pajares, 2006, 2009). In addition, longitudinal studies have further revealed that self-efficacy contributes to academic achievement (e.g., Caprara et al., 2008; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014).

However, the findings from these studies are limited due to their correlational nature. Though valuable, these findings do not demonstrate a *causal* relationship, wherein stronger self-efficacy produces greater academic achievement. According to the research standards set by the What Works Clearinghouse (Institute for Education Sciences, 2014), only experimental research using a randomized controlled trial (RCT) procedure provides evidence that an educational treatment *causes* academic improvement.



This potential causal relationship has been described as a “chicken-and-egg” problem, yet also one that is not contentious because of the reciprocal influence between motivation and behavior (Pajares & Schunk, 2001; see also Bandura, 1986). We disagree. Understanding causality is crucial, in part because it has implications for educational practitioners’ practices. The first author of this article, for example, has been a schoolteacher for nearly 30 years and notes that many of his colleagues believe intuitively that high student self-efficacy evokes desirable learning behaviors and boosts academic achievement. These teachers might be motivated to encourage students to boost their self-efficacy. But the teachers’ efforts are justified only if their intuition is correct.

What is the experimental evidence, then, for a causal relationship between self-efficacy and academic achievement? Surprisingly, the answer at present seems to be: little to none. We searched the literature and identified only a handful of studies that have used various manipulations to successfully alter self-efficacy. But in these studies, changes to self-efficacy do not appear to reliably affect academic achievement (Bouffard-Bouchard, 1990; Jacobs, Prentice-Dunn, & Rogers, 1984; Litt, 1988; Prussia & Kinicki, 1996; Weinberg, Gould, & Jackson, 1979; see also Bandura, 1997, pp. 58–59).

In one of the most cited of these experiments, for example, college students were given positive or negative feedback about their performance on a verbal concept formation task, irrespective of their actual performance. As expected, this feedback changed the students’ self-efficacy: They believed more strongly in their ability to succeed on an upcoming task. But the results also showed that this belief change was unwarranted: The students performed no better than each other on later tasks (Bouffard-Bouchard, 1990).

Other studies have shown that a variety of manipulations—rewards, goal setting, modeling, feedback, task strategies, self-monitoring, self-evaluation, and assessment—enhance students’ self-efficacy (Schunk, 1982; Schunk, Hanson, & Cox, 1987; Schunk & Swartz, 1993; van Dinther et al., 2011). In a recent example, researchers examined how the coursework for pre-service primary teachers influenced their classroom management self-efficacy. Undergraduates in a four-year teacher education program for primary education learned various teaching skills, including classroom management strategies. This coursework elevated their self-efficacy for classroom management (O’Neill, 2016). Unfortunately, this study illustrates the “chicken-and-egg” problem in investigations of self-efficacy as a cause of learning: Behavioral changes—learning of effective task-specific strategies—were first necessary to induce improved self-efficacy. The conclusion that improved self-efficacy is a *result* rather than a cause is unlikely to be attractive to most teachers, who already build

student self-efficacy when teaching new knowledge and skills. If these efforts do not cause improvements in academic achievement, the teachers’ efforts may be better spent elsewhere.

Mori and Uchida (2009) created a unique procedure to test the extent to which improved self-efficacy promotes academic achievement. In their procedure, they used equipment consisting of two distinct images projected on a single screen. Each projected image is viewable only with an appropriate pair of polarizing glasses. They used this equipment to secretly present two different series of anagram tasks to students, such that one group saw easier anagram tasks than their classmates. This “easy” group solved more anagrams and as a consequence reported greater self-efficacy—measured as how well they believed they could perform on the anagram task. This procedure has a number of experimental and practical strengths. First, students can be randomly assigned to conditions. Second, it boosts self-efficacy directly, without relying on prior training of a separate skill. Third, it can be used easily in classroom settings.

Unfortunately, the Mori and Uchida (2009) experiment included only 24 target participants. This small sample size made it difficult to determine the key effect of interest—changes in academic achievement—with any degree of precision. The present research solves this problem by replicating the study with a larger sample. We conducted an experiment using three annual cohorts (comprising six classes each year) from the seventh grade of a junior high school, for a total of 315 participants. We hypothesized that an induced successful performance would promote students’ self-efficacy and, ultimately, their academic performance. We registered this study on the Open Science Framework’s (OSF) website (registration ID: 10.17605/OSF.IO/54WM7) as a replication study with a larger sample: <https://osf.io/54wm7/>.

Method

Participants

We recruited seventh-grade junior high school students from six classes each year for three years from a municipal school in Japan, giving us an initial pool of 656 students (approximately 220 students each year). Twenty-five of these students were absent from the pre-assessment of self-efficacy and were therefore not part of the study, reducing the initial sample to 631 (335 males and 296 females). The socioeconomic status of the students’ families varied within a narrow middle-class range. All students were Japanese natives. The students ranged in age from 12 to 13 years old.

A small number of students were absent on the day of the anagram task ($n = 9$). Because the anagram task was crucial, we excluded these students. In addition, some students were absent for one or more of the repeated assessment periods during the study ($n = 92$). We excluded these students too. In an effort to avoid floor and ceiling effects, only those

students who were within the 26–75 percentile range of scholastic achievement were assigned to our experimental conditions ($n = 315$; $n = 267$ after the exclusions listed above). The remaining 307 non-experimental students participated in the anagram task, but only to ensure consistency of classroom activity—they were not considered part of the experiment proper. For transparency and clarity, we have prepared an anonymized raw data file. This file is available on the OSF at <https://osf.io/cp8uh/>.

Experimental Design

We used a factorial design with two between-subjects factors: treatment group (success, control) and gender (male, female). We included gender in the design because previous literature sometimes finds gender differences in self-efficacy. One study, for example, found that females reported higher self-efficacy in languages and arts than males, while males reported higher self-efficacy than females in mathematics and sciences (Huang, 2013). Another study found that females reported lower self-efficacy than males for a computerized science education task (Nietfeld, Shores, & Hoffmann, 2014). However, other studies have found no gender differences (Caprara et al., 2008; Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011; Jacob, Lanza, Osgood, Eccles, & Wigfield, 2002; Murayama et al., 2013).

Success and Control Students. We randomly selected four to six students in each of the six seventh-grade classes in each of the three year-cohorts as targets to experience success. The remaining students formed the control group. This sampling procedure produced a total of 84 success students (41 males, 43 females) and 231 control students (116 males, 115 females) respectively. We chose this sample size to achieve a statistical power of .8 for detecting a small to medium difference ($d = .4$) between the academic performance of the two groups (Cohen, 1988). The experiment required that only a small number of students in each class experience “success” in order to seem impressive and promote self-efficacy. We therefore limited the number of successful students in each class to between four and six students. Accordingly, there were fewer students in the experimental condition than in the control condition. Because of the nested nature of the sampling procedure, we ran an ANOVA on pre-experimental achievement scores across the 18 classes to examine the influence of class cohorts. We found no meaningful differences, $F_{(17, 638)} = .78$. We also ran an ANOVA on anagram task scores and found no meaningful differences, $F_{(17, 614)} = .56$.

Dependent Variables

We repeatedly assessed two dependent variables: academic achievement and self-efficacy. We operationalized academic achievement as the scores from officially administered school

examinations. We operationalized self-efficacy as students’ self-reports of their ability to complete the anagram task. Details of these assessment procedures are as follows.

Academic Achievement. The junior high school provided us with Z-scores of students’ scholastic achievement. These Z-scores are commonly used in Japanese junior high schools. The scores are standardized and converted such that the mean of the distribution becomes 50 and the standard deviation 10 (Mori & Uchida, 2012). The Z-scores were calculated from the combined scores of term examinations in five major school subjects: Japanese language, social studies, mathematics, natural sciences, and English language. We obtained these Z-scores at six of the school’s assessment periods: prior to the experiment, and then two, five, 10, 14, and 17 months afterward.

Self-Efficacy. We defined self-efficacy procedurally in this study as a student’s rating in response to this specific question: “How well can you perform in the letter rearrangement game?” Students indicated their answer on a five-point scale, ranging from 1 (very badly) to 5 (very well). We assessed self-efficacy eight times (pre-test, post-test, and at six follow-ups). The self-efficacy question was printed on a sheet mixed with other filler questions to mask the experiment’s purpose. As a cover story for administering the questionnaire repeatedly, we told students we were regularly assessing their study habits. The same self-efficacy questionnaire was used in each assessment.

Experimental Procedure

Anagram Tasks. The anagram task was a one-time experience for each student. We ran student participants in class groups. Homeroom teachers led their class—approximately 35–40 students—to a room specially set up for the experiment at the junior high school. We arranged the seats in the room in front of a rear projection screen (80 cm × 80 cm). Students sat in the same configuration as they would in their ordinary classroom (See Figure 1). We prepared two types of polarizing sunglasses beforehand; four to six pairs of one type for the success students and the remaining pairs of the other type for the rest. We placed a pair of polarizing sunglasses on each seat, but only the success students wore the special polarizing sunglasses that let them alone view the easier anagram tasks. To the students, all the sunglasses looked identical. As a cover story, we told the students that the sunglasses were to eliminate glare from the rear projection apparatus.

After the students sat down and put on the sunglasses, the experimenter gave general instructions. Then, he handed an answer sheet to each participant. Next, he projected 30 anagram tasks one-by-one using a PowerPoint slide show on an Apple iBook.

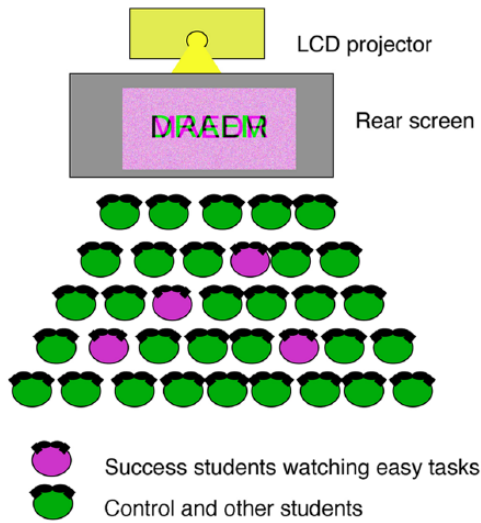


FIGURE 1. Diagram of the experimental setting.

Each of the 30 anagram tasks consisted of five Japanese *hiragana* characters. We arranged 10 of these tasks to have two levels of difficulty in accord with the student's condition (e.g., students in the success condition saw the relatively easy "DRAEM," while subjects in the control condition saw the relatively difficult "MAEDR," both of which can be rearranged to "DREAM"). The remaining 20 anagram tasks had a single problem and solution. We projected the anagram tasks using dual overlapping projections onto a single screen, as depicted in Figure 1. For the 10 tasks with two levels of difficulty, students saw only one version through the polarizing sunglasses (for details of this presentation trick, see Mori, 2007).

We presented each anagram task for 10 seconds. During this time, the students tried to solve each anagram and write the answer on their answer sheet. We also included a five-second interval between each anagram task. The experimenter asked the students to stop writing at the end of the anagram task.

Next, the experimenter announced the correct answers so that students could mark their answers. Then, the experimenter asked students with more than 22 correct answers to raise their hands. These students were frequently met with spontaneous applause from the class. Because no students were aware of the presentation trick, we assume these naturally occurring appraisals were genuine. We did not anticipate nor control for applause, and therefore did not collect data concerning any potential effects of applause in this study.

Debriefing. Approximately one month after the anagram task, we disclosed the experimental purpose and the sunglasses trick to the students. But we did not specify which

students, specifically, had observed the easier versions of the anagrams.

Results

Manipulation Check

We first examined whether students who viewed easier anagrams solved more anagrams correctly. As expected, the success students answered more anagram tasks correctly ($M = 24.90$, $SD = 3.90$, range = 3–29) than the control students ($M = 20.04$, $SD = 3.49$, range = 5–29). We also found that males answered fewer anagram tasks correctly ($M = 21.31$, $SD = 4.49$, range = 3–29) than female ($M = 22.14$, $SD = 3.68$, range = 11–29). A 2 (treatment group: success, control) \times 2 (gender: male, female) ANOVA revealed a statistically significant effect of treatment group: $F_{(1,309)} = 114.09$, $p < .001$, Cohen's $\eta^2 = .26$, and a statistically significant effect of gender: $F_{(1,309)} = 8.54$, $p = .004$, $\eta^2 = .02$. The interaction was not statistically significant, $F_{(1,309)} = .92$, $p > .250$.

Upon closer examination of the data, we noted that 10 of the success students scored fewer than 22 correct answers on their easier version of the anagram task, while 80 of the control students scored 22 or more correct answers on their harder version. These scores are incongruent with the experimental manipulation. However, because we found that the pattern of results remained virtually unchanged when these subjects were excluded, we elected to include these students in our analyses.

Some students were absent from one or more occasions of the self-efficacy assessments and the academic achievement tests. We followed the same process as in a previous study, deleting these missing data case-wise (Mori & Uchida, 2009). Case-wise deletion procedures have at least two strengths: (a) for education RCTs that focus on test score outcomes, case deletion performs reasonably well relative to other missing data adjustment methods; and (b) case deletion is simple to apply and understand (Schochet, 2016; pp. 53–54). Ultimately, there were 267 students with complete assessment data for the following analyses (72 in the experimental condition and 195 in the control condition). For transparency, the raw data file with all data from students who participated in the study is available on the OSF site: <https://osf.io/cp8uh/>.

Self-Efficacy

We assessed students' self-efficacy at eight periods; these data appear in Figure 2. As the figure shows, students' self-reports of their ability to perform well on the anagram task rose sharply after the anagram task and remained high for one year—but only for those students in the success condition. The control students' self-efficacy, on the other hand, remained virtually unchanged. A 2 (treatment group: success, control) \times 2 (gender: male, female) \times 8 (assessment period) mixed ANOVA

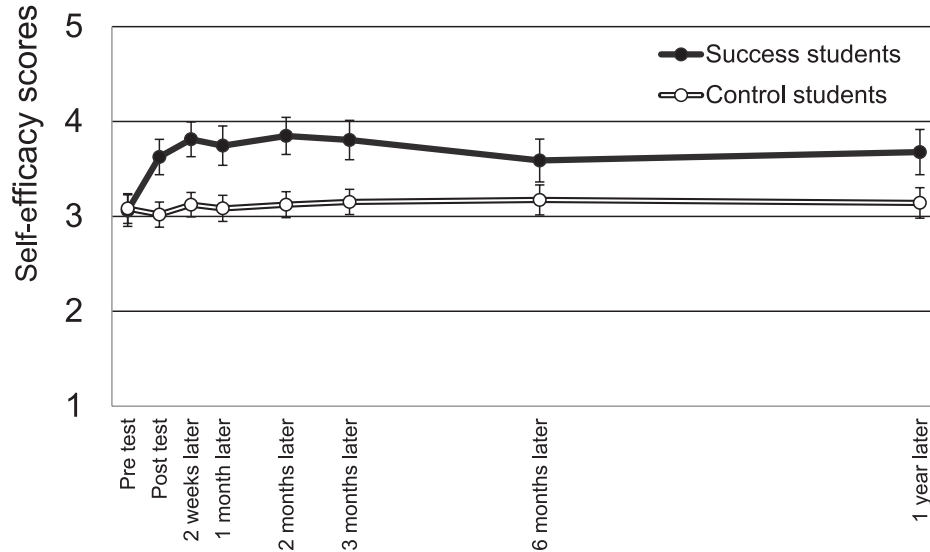


FIGURE 2. Self-efficacy scores before and after the anagram task. The vertical bars indicate the 95% confidence intervals.

revealed an interaction between treatment group and assessment period; $F_{(7,1841)} = 10.95, p < .001, \eta^2 = .03$. Follow-up comparisons using the Ryan procedure (Ryan-Einot-Gabriel-Welsch and Quot [REGWQ] procedure) showed that success students reported greater self-efficacy than control students at all assessment periods, except before the anagram tasks: $F_s > 12.60, p_s < .0004$. We found no statistically significant main effects nor interactions with gender: $F_s < 1.30, p_s > .255$.

Academic Achievement

We obtained students' average Z-scores at each of six assessment periods, including before the experiment, and then two, five, 10, 14, and 17 months afterward; these data appear in Figure 3, split by gender. As the figure shows, the Z-scores of males in the success condition increased from pre-test at the two-month assessment period, and remained elevated. In contrast, the Z-scores of males in the control condition showed a declining tendency.¹ For females, we found no clear differences between the two experimental conditions.

A 2 (treatment group: success, control) \times 2 (gender: male, female) \times 6 (assessment period) mixed ANOVA revealed a statistically significant treatment group \times gender \times assessment period interaction, $F_{(5,1315)} = 2.69, p = .020, \eta^2 = .01$; and a gender \times assessment period interaction, $F_{(5,1315)} = 3.14, p = .008, \eta^2 = .01$. We found no other statistically significant main effects or interactions ($F_s < 3.23, p_s > .071$). To unpack the three-way interaction, we performed a 2 (treatment group: success, control) \times 6 (assessment period) mixed ANOVA for the males, and another for the females. For the males, we found a significant interaction ($F_{(5,630)} = 4.41,$

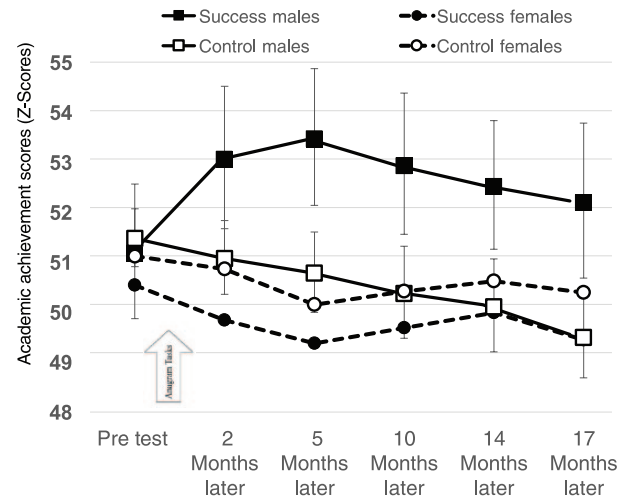


FIGURE 3. Academic achievement scores before and after the anagram task. The vertical bars indicating the 95% confidence intervals are added only to males' data.

$p < .001, \eta^2 = .03$), revealing that the differences between Z-scores of success and control males changed over the assessment periods; in general, the Z-scores of success males increased after the anagram task ($F_{(5,630)} = 4.26, MS = 30.60, MSe = 7.18, p < .001$) while those of the control males declined ($F_{(5,630)} = 3.24, MS = 23.20, MSe = 7.18, p = .007$). Multiple comparisons by the Ryan procedure (REGWQ) showed statistically significant greater Z-scores for success males at five months after the self-efficacy manipulation, compared to their pre-experiment scores. Meanwhile, the

Z-scores of the control males declined gradually and reached statistically significant differences from pre-experimental scores at 10, 14, and 17 months after the experiment. For the female students, however, we found no statistically significant effects ($F_s < 2.06$, $p_s > .069$).

A cautious reader may wonder about potential problems that arise due to the nested nature of the data, such as deteriorating statistical power (Usami, 2013, 2014). But note that our study used multisite randomization trials (MRT). That is, students were randomly assigned to experimental and control conditions in each class. Simulations show that MRT procedures produce relatively stable intra-class correlations when compared with clustered randomization trials (CRT) (see Table 1a and 2a in Usami, 2011). Nonetheless, we tested the effect of year-cohort differences by including year-cohort as a variable in a three-way ANOVA for the males ($3 \text{ year-cohorts} \times 2 \text{ treatment groups} \times 6 \text{ assessment periods}$). This analysis revealed a statistically significant interaction for treatment group \times assessment period ($F_{(5,610)} = 3.93$, $p < .01$, $\eta^2 = .006$) and a significant main effect for assessment period ($F_{(5,610)} = 2.67$, $p < .05$, $\eta^2 = .004$). All other effects failed to reach statistical significance ($F_{(2,122)} = 0.24$ for the main effect of year-cohort, $F_s = 0.84$ and 1.15 for the interactions). Consistent with this analysis, we also found a similar pattern displayed in Figure 3 when we looked separately at each of the three cohorts (see supplemental figures on the OSF site (<https://osf.io/kuerw/>)).

Discussion

Experimental Enactment of Self-Regulated Learning

Across three annual cohorts comprising a total of 267 students, we found in an RCT experiment that a brief experience of success in an anagram task raised students' self-efficacy immediately and eventually improved the male students' overall academic performance. Moreover, these broad benefits remained more than one year after the brief experimental manipulation.

How does an induced successful experience in a simple task lead to overall academic improvement? One possible answer comes from self-regulated learning theory (Zimmerman, 1990), which hypothesizes a "virtuous causal cycle": Students first experience success, which raises their self-efficacy. Improved self-efficacy increases motivation and the use of effective learning strategies. These covert and overt changes then lead to improved academic achievement, and the cycle begins anew.

Previous researchers have examined and found support for this self-regulated learning hypothesis using correlational methods (Caprara et al., 2011; Chen & Usher, 2013; Murayama et al., 2013; Usher & Pajares, 2009; Zuffianò et al., 2013). In our study, we provide novel experimental evidence in support of the theory. The "virtuous cycle" began with contrived success on a specific task, which

strengthened self-efficacy for that task. But intriguingly, the benefits extended beyond the task itself to students' broad academic ability.

These results may be particularly encouraging to teachers looking to break their students out of a cycle of poor scholastic performance and low self-efficacy. Here, we present some evidence that, to break that cycle, teachers might give these students "easy" tasks so that they experience success. We enclose the word easy in quotation marks because, according to what we have demonstrated here, the tasks should be covertly easy only for target students. We hope schoolteachers will come up with creative ways to accomplish this requirement.

Gender Differences

We found that boosting self-efficacy improved academic scores, but only among males and not females. Why? As briefly described earlier, the literature is mixed with respect to gender differences in academic self-efficacy. Some studies find differences (Huang, 2013; Nietfeld et al., 2014), but others do not (Caprara et al., 2008, 2011; Jacob et al., 2002; Murayama et al., 2013). We were unable to find a good explanation within these studies that could account for our pattern of results.

Instead, one possible explanation relates to the different attributions males and females make about their ability to succeed. Males, for example, will more readily ascribe the cause of their success to ability than females (Lloyd, Walsh, & Yailagh, 2005). Perhaps, then, the males in our study were more likely than females to attribute success on the anagram task to their ability, while females were more likely than males to attribute success to luck. If true, then an ability-based attribution may be necessary to see general academic improvements. We state this potential explanation cautiously, however, because we found that males and females both reported increased self-efficacy, and we did not measure students' attributions of success. It may be worthwhile to ask participants about these attributions in future research.

Feedback Effects

The students in the success condition might have raised their self-efficacy and achievement scores simply because they received positive feedback. However, a review of the feedback literature, and other studies collected through major educational databases, concluded that there were inconsistent findings with respect to feedback (Shute, 2008). Some findings reported no feedback effects (Sleeman, Kelly, Martinak, Ward, & Moore, 1989) or even negative effects on learning (Kulhavy, White, Topp, Chan, & Adams, 1985). The review ultimately concluded that feedback could improve learning processes and outcomes, but only under certain conditions (Shute, 2008).

Feedback effects have been inconsistent because there are a variety of intervening variables (Krenn, Wuerth, and Hergovich, 2013). Moreover, self-efficacy is one of these moderating variables. Managers with high self-efficacy, for example, benefit more from feedback than those with low self-efficacy (Heslin & Latham, 2004). Feedback might also have effects on learners' attitudes and beliefs. For example, feedback attributed to competence promoted self-efficacy more in third-grade children on a subtraction skill test than feedback attributed to effort (Schunk, 1983). Considered as a whole, these findings bring us back to the "chicken-and-egg" problem of causal relations. We believe, therefore, that there is insufficient evidence to attribute the promotion of self-efficacy of students in our study merely to positive feedback.

Limitations and Directions for Future Research

The most crucial limitation of our study is that it is unclear how the initial experience of success produced greater academic achievement. Although our hypothesis was theoretically motivated—drawn from the literature on self-regulated learning—the proximal and distal mechanisms were not well specified (Zimmerman, 1990). There are, potentially, a number of intervening variables, including: attributions of success, gradual transformations of task-specific to general self-efficacy, and increased motivation. A more complete explanation of the effects we report here will likely require future assessment of these variables, to untangle their contribution. Such work will illuminate the processes intervening between initial success and later achievement.

We have demonstrated a brief intervention for students, producing remarkable results. Teachers—who want to motivate students, especially those with low confidence or learning difficulties—may wish to capitalize on this intervention. But there are limitations in applying this intervention to actual school settings. First, only a small number of students can experience success, or it would cease to be remarkable. Educators are unlikely to want to adopt the practice if it can be used only for a fraction of students. Second, and relatedly, our study used minor deception, revealing the "trick" to students only at debriefing. This necessary deception is also likely to make it difficult for educators to adopt the practice.

We also note that our students' experiences of success included appraisal from classmates in the form of applause. This unintentional social appraisal occurred naturally and was thus outside our control. It is therefore more appropriate to regard the induced success we used here as induced success with social appraisal. At present, we do not know how this social appraisal affects student behavior. A follow-up experiment controlling for the presence or absence of appraisal could usefully tease apart the influence of an induced experience of success from the influence of appraisal.

Finally, the gender differences in academic achievement pose an intriguing research question ripe for future investigation. Although our explanation above for this difference is plausible, we currently have no direct or even indirect evidence to suggest it is true. We plan to address this issue in a future experiment that probes students' attributions of success.

Conclusions

Our study provides a real-world experimental enactment of Bandura's self-regulatory efficacy theory in junior high school students. We hypothesized that a single experience of success would promote students' self-efficacy. Using a presentation trick, we secretly presented easier anagram tasks to target students, which led to an experience of success. These success-induced students reported improved self-efficacy and maintained this improved self-efficacy over an entire year. Most importantly, the success-induced males showed significant improvement in their academic achievement. It is unclear, at present, why improved self-efficacy produced higher achievement only in males. Nonetheless, our findings may give hope to teachers seeking a means to encourage students who suffer from low self-efficacy.

Acknowledgments

This research was supported by KAKENHI from JSPS to AU (Grant No.19905005) and to KM (Grant No.16653054 & 23653182). It was approved by the ethical committee of Tokyo University of Agriculture and Technology. We thank the teachers and pupils of Shinonoi Nishi Junior High School for participating in this research. We are indebted to Kou Murayama and Magdalena Abel for their comments on an earlier draft. We are also grateful for Satoshi Usami for providing appropriate advice on statistical analyses. Special thanks are due to Maryanne Garry for her support during the manuscript writing.

Note

1. This decline was illusory. The initial average Z-scores of the target and control groups selected randomly from the 26–75 percentile ranges were .5–1.5 points above 50. That was because the distribution skewed leftward. Mori and Uchida (2012) found the leftward skews tended to occur at the beginning of the school year and become less skewed gradually after that.

ORCID iDs

Robert B. Michael  <https://orcid.org/0000-0001-5275-7636>

Kazuo Mori  <https://orcid.org/0000-0002-9631-5845>

References

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology, 130*, 353–363.
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*, 485–499.
- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*, 525–534.
- Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology, 81*, 78–96.
- Chen, J. A., & Usher, E. L. (2013). Profiles of the sources of science self-efficacy. *Learning and Individual Differences, 24*, 11–21.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Heslin, P. A., & Latham, G. P. (2004). The effect of upward feedback on managerial behavior. *Applied Psychology, 53*, 23–37.
- Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*, 1–35.
- Institute for Education Sciences. (2014). *What works clearing-house procedures and standards handbook (Version 3.0)*. Washington, DC: Institute for Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf
- Jacob, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grade one through grade twelve. *Child Development, 73*, 509–527.
- Jacobs, B., Prentice-Dunn, S., & Rogers, R. W. (1984). Understanding persistence: An interface of control theory and self-efficacy theory. *Basic and Applied Social Psychology, 5*, 333–347.
- Krenn, B., Wuerth, S., & Hergovich, A. (2013). Individual differences concerning the impact of feedback – Specifying the role of core self-evaluations. *Studia Psychologica, 55*, 95–110.
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*, 285–291.
- Kupermintz, H. (2002). Affective and conative factors as aptitude resources in high school science achievement. *Educational Assessment, 8*, 123–137.
- Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8*, 139–162.
- Litt, M. D. (1988). Cognitive mediators of stressful experience: Self-efficacy and perceived control. *Cognitive Therapy and Research, 12*, 241–260.
- Lloyd, J. E. V., Walsh, J., & Yailagh, M. S. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know it? *Canadian Journal of Education, 28*, 384–408.
- Mori, K. (2007). Projecting two words with one machine: Presenting two different visual stimuli using just one projector without viewers' noticing the duality. *Behavior Research Methods, 39*, 811–815.
- Mori, K., & Uchida, A. (2009). Can contrived success affect self-efficacy among junior high school students? *Research in Education, 82*, 60–68.
- Mori, K., & Uchida, A. (2012). The leading group effect: Illusionary declines in scholastic standard scores of mid-range Japanese junior high school students. *Research in Education, 87*, 64–74.
- Murayama, K., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2013). Predicting long-term growth in adolescents' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84*, 1475–1490.
- Nietfeld, J. L., Shores, L. R., & Hoffmann, K. F. (2014). Self-regulation and gender within a game-based learning environment. *Journal of Educational Psychology, 106*, 961–973.
- O'Neill, S. (2016). Preparing preservice teachers for inclusive classrooms: Does completing coursework on managing challenging behaviours increase their classroom management sense of efficacy? *Australasian Journal of Special Education, 40*(2), 117–140. doi:10.1017/jse.2015.10
- Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology, 25*, 406–422.
- Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. Riding & S. Rayner (Eds.), *Self-perception* (pp. 239–266). Westport, CT: Ablex Publishing.
- Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*(1), 29–48.
- Prussia, G. E., & Kinicki, A. J. (1996). A motivational investigation of group effectiveness using social-cognitive theory. *Journal of Applied Psychology, 81*, 187–198.
- Schochet, P. Z. (2016). *Statistical theory for the "RCT-YES" software: Design-based causal inference for RCTs. NCEE 2015-4011 (Second Edition)*. National Center for Education Evaluation and Regional Assistance. Retrieved from <https://files.eric.ed.gov/fulltext/ED556496.pdf>
- Schunk, D. H. (1982). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology, 74*, 548–556.
- Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*, 848–856.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*, 54–61.
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*(3), 337–354.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.
- Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science, 13*, 551–568.

- Usami, S. (2011). A unified method for determining sample size needed to evaluate mean difference in hierarchical research design and construction of numerical table—Focusing on statistical power and confidence interval of effect size. *Japanese Journal of Educational Psychology*, 59, 385–401.
- Usami, S. (2013). Generalized sample size determination formulas for experimental research with hierarchical data. *Behavior Research Methods*, 46, 346–356.
- Usami, S. (2014). A convenient method and numerical tables for sample size determination in longitudinal-experimental research using multilevel models. *Behavior Research Methods*, 46, 1207–1219.
- Usher, E. L., & Pajares, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology*, 31, 125–141.
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34, 89–101.
- van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6, 95–108.
- Weinberg, R., Gould, D., & Jackson, A. (1979). Expectations and performance: An empirical test of Bandura's self-efficacy theory. *Journal of Sport and Exercise Psychology*, 1, 320–331.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25, 3–17.
- Zuffianò, A., Alessandri, G., Gerbino, M., Kanacri, B. P. L., Di Giunta, L., Milioni, M., & Caprara, G. V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learning and Individual Differences*, 23, 158–162.

Authors

AKITOSHI UCHIDA is a schoolteacher in mathematics at Sairyo Junior High School, Nagano, Japan. His research interests are scientific examinations of mathematics education.

ROBERT B. MICHAEL is an assistant professor at the Department of Psychology, the University of Louisiana at Lafayette, in the United States. He has been doing memory research for applied purposes.

KAZUO MORI currently works at the Department of School Education, Matsumoto University, Matsumoto, Japan. His research interests include experimental demonstrations and examinations of various issues.