

Sources of Self-Efficacy Belief: Development and Validation of Two Scales

Ou Lydia Liu
Educational Testing Service, Princeton

Mark Wilson
University of California, Berkeley

Self-efficacy belief has been an instrumental affective factor in predicting student behavior and achievement in academic settings. Although there is abundant literature on efficacy belief per se, the sources of efficacy belief have not been fully researched. Very few instruments exist to quantify the sources of efficacy-beliefs. To fill this void, we developed two scales for the two main sources of self-efficacy belief: *past performance* and *social persuasion*. Pilot test data were collected from 255 middle school students. A self-efficacy measure was also administered to the students as a criterion measure. The Rasch rating scale model was used to analyze the data. Information on item fit, item design, content validity, external validity, internal consistency, and person separation reliability was examined. The two scales displayed satisfactory psychometric properties. Applications and limitations of these two scales are also discussed.

Introduction

With the publication of "Self-efficacy: Toward a Unifying Theory of Behavioral Change," Bandura (1977) identified an important element in cognitive psychology: self-efficacy belief. This concept is defined as an individual's belief in his or her own capability to organize and execute the actions required to achieve specific goals (Bandura, 1977). Individuals create and develop self-perceptions of capability, which in return help determine their behavior and persistence in pursuing desired goals.

According to Bandura (1977), a few sources contribute to the development of self-efficacy beliefs. Past performance and social persuasion are two very important sources. Past performance has been found to be the most effective way of creating a strong sense of efficacy. Constant successes can help build and sustain a strong belief in one's personal efficacy, while constant failures undermine it, especially when failures occur prior to the establishment of a robust efficacy. Previous research has found that past performance has the most influential effect on subsequent efficacy beliefs (Feltz, 1994, 1988; McAuley, 1985).

Social persuasion is another important way of strengthening individuals' beliefs that they have the required abilities to succeed in certain tasks. People who are reminded or persuaded that they possess the capabilities to perform certain tasks are more likely to exert and sustain extra effort than if they were reminded to focus on their personal deficiencies. However, verbal persuasion must be realistic, and the strength of the influence depends on the content of the message delivered to the recipients of the verbal persuasion (Chase, 1998). If the verbal persuasion recipients realize that they cannot achieve the things they are encouraged to do, then the persuasion might actually have no or even a negative effect on their self-efficacy belief.

According to Bandura (1977), how people behave can often be better predicted by the beliefs they hold about their own capabilities than by their actual achievement. These self-perceptions, called self-efficacy beliefs, help determine the

way in which individuals act using the knowledge and skills they have. Based on the proliferation of research over the past 30 years, self-efficacy belief has been identified to affect people's behaviors in a wide range of areas. Self-efficacy belief was found to mediate athletic persistence, choice of college major, career choice, quitting smoking efforts, and young drinker's alcohol consumption (Bandura, 1986; Donnay and Borgen, 1999; Lucas, Wanberg, and Zytowski, 1997; Ramalingam and Wiedenbeck, 1998; Smith, 2001). An abundance of research has been devoted to the predictive power of self-efficacy belief in academic performance. Results revealed that efficacy belief has a strong influence on various academic subjects such as reading, writing, math, science, and computer achievement (Lent, Lopez and Bieschke, 1991; Schunk and Rice, 1993; Pajares and Johnson, 1996; Pajares and Kranzler, 1995; Pajares and Miller, 1994, 1995; Shell, Colvin, and Bruning, 1995; Schunk, 1991, 1990).

While it is helpful to realize that efficacy beliefs are powerful predictors for certain human behaviors, it is equally important to understand how efficacy beliefs are developed and the ways in which they mediate the outcomes. Pinpointing the factors that contribute to the establishment of efficacy beliefs would help clarify the sources of efficacy belief, and may benefit remedial programs that aim to promote students' efficacy beliefs.

Compared to the extensive research on how self-efficacy beliefs affect human behavior, the sources of self-efficacy belief have not been investigated with the rigor they deserve, with the exception of a few studies. For example, Anderson and Betz (2001) developed four scales for four sources of efficacy belief (past performance, social persuasion, vicarious experience, and physiological states) and concluded that for mathematics self-efficacy, the impact of past performance, social persuasion, physiological states, and direct learning experiences, is distinct from that of vicarious learning, which is an indirect learning experience. Wise and Trunnell (2001) conducted a study examining the influence of different sources of efficacy information on the strength of self-

efficacy in bench pressing. Results indicated that an accomplishment led to significantly stronger bench-pressing efficacy than did the observation of a model. While the observation of a model was more effective in strengthening the efficacy belief than receiving a verbal message. A verbal persuasion message was most effective in strengthening efficacy when it was followed by a performance accomplishment.

Smith (2001) examined the influence and interrelationship of four sources of influence on computer self-efficacy. The results indicated that past performance correlated significantly with vicarious experience, social persuasion and physiological states. An examination of gender and ethnic differences revealed that past performance, among all sources, was the most influential source of computer self-efficacy for white males. Further, Richard (2005) noted that previous computer knowledge, such as database software skills, was a contributor to positive computer self-efficacy. Finally, in a study exploring the sources of math self-efficacy expectations and their relationship with math self-efficacy, Ozyurek (2005) reported that the measurement of past accomplishments significantly predicted math self-efficacy while vicarious experience did not.

Most of the studies on sources of efficacy beliefs target the beyond-adolescence population, and not enough research has been dedicated to exploring the sources of efficacy belief among middle school students. There is evidence that the academic achievement of young learners in the elementary grades, K-6, can be significantly affected by their self-efficacy belief (Jinks and Morgan, 1999; Kim and Lorsch, 2005). Research has also noted that students' (aged 11-15 years) perceived academic and social efficacies directly influence the types of career activities in which they judge themselves to be efficacious (Bandura, Claudio, Gian, and Concetta, 2001). According to Bandura (1986), self-efficacy belief was not conceived as a developmental theory, as the sources of self-efficacy belief are not expected to change with age and cognitive proficiency levels. Negative impact resulting from low efficacy belief in childhood and adolescence may linger to adulthood.

Furthermore, efficacy beliefs are strongly related to other motivational constructs such as confidence, self-concept, and self-esteem (Betz and Klein, 1996; Paulsen and Betz, 2004), which play instrumental roles in adulthood success. Therefore, it is of vital importance for middle school students to establish and maintain high levels of efficacy beliefs for the benefit of their further educational and occupational success. To meet this end, this study developed two scales to measure student past performance and social persuasion. The psychometric properties of these scales were investigated in terms of validity, reliability, and correlation to a criterion measure. It is hoped that these scales will provide useful information on how students build and sustain their academic self-efficacy and the major factors that affect the establishment of a strong sense of efficacy beliefs.

Instrument Development

Two scales were developed for the two sources of efficacy belief: past performance and social persuasion, with a focus on middle school students. The efficacy beliefs of middle school students are less established than those of older students, thus more malleable.

Past performance was selected for investigation because it has been reported by many studies to be the most influential source for efficacy beliefs (Feltz, 1994, 1988; Feltz and Riessinger, 1990; McAuley, 1985; Ozyurek, 2005; Richard, 2005; Weinberg, Grove, and Jackson, 1992). Social persuasion was selected because it has been the most widely used source in school settings due to its easy availability. Evidence about how past performance and social persuasion predict self-efficacy belief can provide valuable information for the activities and programs that aim to boost student efficacy belief.

An initial version of 20 items was developed for each scale. Students were asked to respond to a 4-point Likert scale (*Strongly Agree, Agree, Disagree, and Strongly Disagree*) to indicate the extent to which each statement characterized them. Previous research showed that compared to the 10-point Likert type scale, the 4-point Likert

type scale is more parsimonious and has higher reliability (Smith, Wakely, Kruij, and Swartz, 2003).

Self-report items on the past performance scale were designed to collect individuals' appraisal of their experiences concerning past academic successes and failures. Information was elicited from students on their general academic performance (i.e., I get high scores in most exams) and performance in specific subjects (i.e., I have been doing very well reading). These items were selected after consulting with a panel of educational researchers, teachers, parents, and students, who indicated that these items comprised a major part of students' academic life.

Items on the social persuasion scale also relied on self-reports to assess the student's memory on the kinds of social persuasion they received from parents, teachers, and peers. The social persuasion scale measures the nature of the comments and feedback the students received on academic performance. Some of these items describe general comments (i.e., "If I fail at something, my parents encourage me to try again"), while others target specific subjects (i.e., "My teachers tell me I have low math ability"). Before finalizing the decision of focusing on the aforementioned types of comments for the social persuasion scale, opinions from a group of teachers, parents, and middle school students were once again solicited.

Items on the past performance scale and the social persuasion scale were reviewed by a panel, consisting of experts in psychological measurement, professors in motivational studies, researchers in the general educational field, and classroom teachers. The review and discussion focused on item content, item representation of the construct, reading level, relevance to student daily life, and appropriate level of privacy. A small group of 6th - 8th grade students ($N = 20$) from two public schools in the bay area of California participated in a pilot test, aimed at collecting information on overall clarity and accessibility of the items. Preliminary results indicated that students were able to understand the items.

After the revisions were made based on the feedback from various sources, 13 items were finalized for the past performance scale and 17 items for the social persuasion scale (See Appendix). The scoring of these items was straightforward. Scores 0 to 3 are assigned to the four categories from strongly disagree (0) to strongly agree (3). The scores were reversely recoded for negatively worded items (i.e., I often fail in exams). After the recoding, a higher score on the past performance scale indicates a more successful past experience, and a higher score on the social persuasion scale indicates that a positive social persuasion has been received.

It is important to examine the psychometric properties of these two scales in order for them to be used in future studies. Informed by Messick (1989) and Wilson (2005), the concept of validity was explored based on item fit, internal structure (item design), content validity (rank order correlation), and external validity (relations to other efficacy scales). The reliability information was investigated in terms of internal consistency and person separation reliability.

Sample

Pilot test data were collected from 255 (69% male) students in two public schools in the bay area of California. The participants consisted of 6th ($N = 97$, 49% male), 7th ($N = 83$, 52% male), and 8th ($N = 75$, 54% male) grade students. The mean age was 13.5 with a standard deviation of 1.1.

Procedure

Students completed the past performance scale and the social persuasion scale. In order to collect information on external validity, students were asked to respond to a general self-efficacy scale (Sherer and Maddux, 1982). The general self-efficacy scale consists of 17 items, represented by a 4-point Likert scale. To align the items with the student's level of understanding, some of the items were re-worded or modified. For example, the item "I am a self-reliant person" was modified to "I do things on my own without help." A higher score on the general self-efficacy scale indicates a higher level of overall self-efficacy

belief. Students were informed that there was no right or wrong answer to these items. The response process was anonymous so no individual student could be identified.

Methods

The rating scale model (RSM) (Andrich, 1978; Wright and Masters, 1982) from the Rasch item response modeling family was selected to analyze the data. The RSM specifies the probability that respondent *n* will score *x* on a particular rating scale item *i*. The mathematical form of the RSM is expressed as follows:

$$P_{nis}(\theta) = \frac{\exp \sum_{j=0}^x [\theta_n - (\delta_i + \tau_j)]}{\sum_{r=0}^m \exp \sum_{j=0}^r [\theta_n - (\delta_i + \tau_j)]}$$

(1)

$x = 0, 1, \dots, m,$

where θ_n denotes the latent trait variable. For example, for the past performance scale, θ_n represents the student's position on a continuum of success. δ_i is the item difficulty, indicating how likely this item would be endorsed by the respondents. τ_j is a set of additional threshold parameters which describe the difficulty of scoring *x* instead of *x - 1*. The RSM assumes that the distance between two adjacent threshold parameters is the same across all items. For instance, the 13-item past performance scale was represented by four categories; hence, there were 13*3 item parameters. An alternative model, the Rasch partial credit model, assumes that the distance between two adjacent threshold parameters varies across items, and is often applied to instruments containing constructed-response items. In the case of this paper, where all items are represented by 4-point Likert type categories, the rating scale model was chosen over the partial credit model.

The software program ConQuest (Wu, Adams, and Wilson, 1998) was used to perform the analysis. ConQuest provides a weighted fit mean square (WFMS) statistic as a fit statistic index for each item, a discrimination index for each item, and a weighted likelihood estimate (WLE)

for each student (Warm, 1989). The WFMS is an indication of the agreement between the measurement model and the observed data. It has an expected value of 1 and can assume values that range from 0 to ∞ (Wright and Linacre, 1994). Values larger than 1 suggest more variance in the data than the model predicted and values less than 1 suggest a lack of stochasticity in the data. Adams and Khoo (1996) considered .75 as a reasonable lower bound and 1.33 as a reasonable upper bound for this fit statistic.

WLE is a point estimate of a student's ability. Average WLE represents the mean ability or attitude level for all the students who choose a certain category. Theoretically, students who choose a more difficult category should have a higher average WLE than those who cannot reach that category. In this case, a higher WLE reflects a more successful past experience or positive social persuasions received.

The discrimination index provided by ConQuest is from the classical test theory (CTT) framework. It indicates the items' discriminating ability in differentiating between the students who display higher levels of traits and lower levels of traits as measured by the instruments. The discrimination index ranges from -1 to 1 and it is expected to be positively high. According to the conventional rule (Organization for Economic Co-operation and Development, 2005), items with a discrimination index lower than .25 are considered to have poor discriminating power.

ConQuest also offers a standard error of measurement (SEM) for each distinct response vector. The SEM indicates the precision of the estimate, which can be used to calculate the person separation reliability index. The person separation reliability indicates how well an instrument separates respondents from a particular sample (Wright and Masters, 1982), and can be formulated as:

$$R_p = 1 - \frac{MSE_p}{\text{var}(\hat{\theta})}$$

(2)

where $\sqrt{\text{var}(\hat{\theta})}$ is the observed standard deviation of the self-efficacy source measures and

MSE_p is the mean squared standard errors of the person estimates, which equals:

$$\frac{\sum_{n=1}^N SEM_n^2}{N}$$

Results

Results presented in this paper were derived from analyses under the rating scale model implemented in ConQuest. Psychometric properties with regard to validity and reliability were examined for the past performance and the social persuasion scales.

Item Fit. Fit statistics help detect discrepancies between the modeled assumptions and the actual data (Wright and Masters, 1982) and can provide evidence for construct validity in a Rasch modeling context (Wright and Stone, 1988). Here, WFMS was used as the fit indicator. The fit statistics for the two scales are provided in Figures 1 and 2.

Among items on the two scales, all but one item showed satisfactory fit statistic. The misfit item (WFMS = 1.36) was the past performance item, "I often fail in exams." The item information is provided in Table 1.

We can see that the mean WLE was in reverse order for those who chose "strongly agree" and "agree." It is expected that those who chose "agree" should have a higher mean WLE than those who chose "strongly agree." However, the opposite pattern appeared on this item. The discrimination index for this item was .07, far below the rule of thumb number .25. Multiple sources of evidence suggested that this item failed to differentiate among students. A possible reason could be that teachers of these students sampled in this study did not use the pass or fail system when grading students. Therefore, this item was not applicable or relevant for most students.

Wright Map. In the Rasch model, the probability of answering an item correctly or endorsing

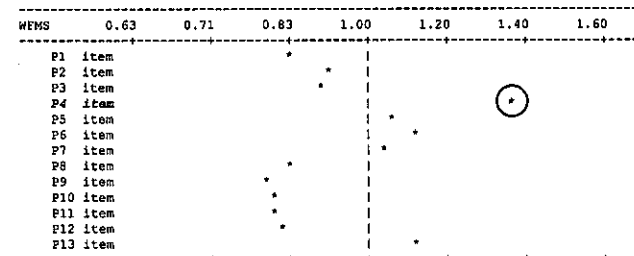


Figure 1. Fit map for the past performance scale.

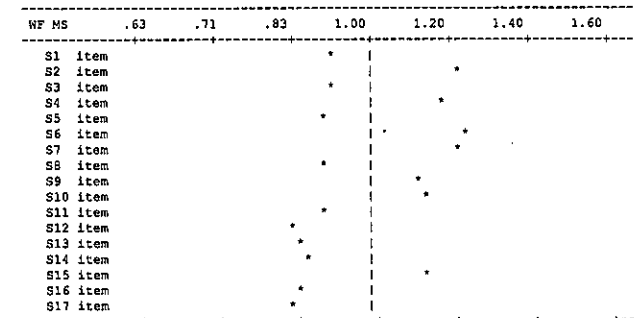


Figure 2. Fit map for the social persuasion scale.

Table 1
Item information for Item P4 on the past performance scale

Score	Category	No. of Students	Percent (%)	Weighted Likelihood Estimate (WLE)
0	Strongly Agree	5	2.3	-.27
1	Agree	34	16.3	-.30
2	Disagree	50	23.9	-.20
3	Strongly Disagree*	74	35.4	.26
Missing		46	22.1	-.62

*Since this is a negatively framed item, the scores were recoded so "strongly disagree" corresponds to the highest score for this item.

the most valued statement is modeled as a function of the respondent ability and item difficulty (Wright and Masters, 1982). A Wright Map is a visual representation of this relationship.

In Figure 3, the "x"s on the left panel represents the students and the numbers in the right panel represent the items and their Thurstonian thresholds. For an item with the maximum score k (0, 1, ..., k), there are k Thurstonian thresholds. The k^{th} threshold can be interpreted as the point at which the probability of a score k and below is equal to the probability of scores above k . All of the items here have four categories so they each have three thresholds. It is expected that, in general, it would be more difficult for students to reach a higher threshold than a lower threshold (e.g., threshold 3 is more difficult to reach than threshold 2). Figure 3 demonstrates that when compared to most other thresholds 1s, the threshold 1 of item P4, is quite difficult to reach, suggesting that it is hard for students to agree with this item statement. As noted earlier, a likely reason was that teachers did not fail students on tests so very few students chose this category. Teachers could give "satisfactory" or "unsatisfactory" labels to student marks instead of "pass" or "fail." Given the irrelevance of this item, it was removed from the past performance scale.

Item Design. Item design represents another important aspect of construct validity (Embretson and Gorin, 2001). Construct validity is evident when high-scoring respondents on the construct also, in general, score higher on each item (Wilson, 2005). If the mean ability estimate of students choosing each item category tends to increase as the category score increases, then it is reasonable

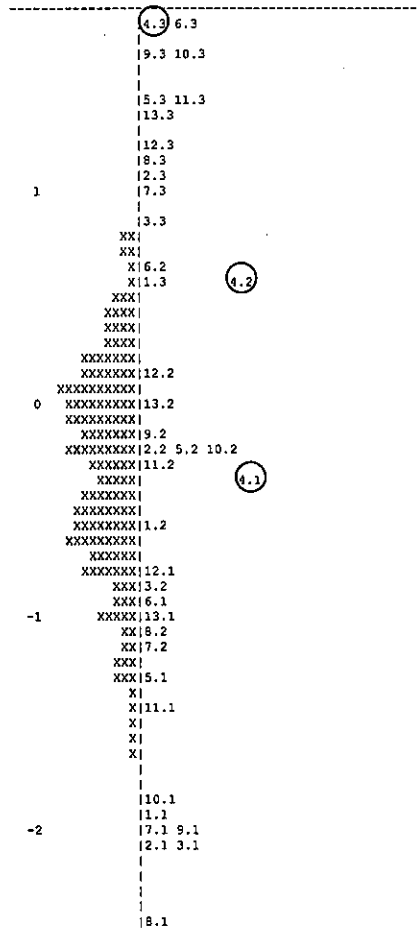


Figure 3. Wright map for the past performance scale.

to say that this particular expectation of item design has been fulfilled.

For items developed in this study, mean estimates should increase from the category "strongly disagree" to "strongly agree" for the positively framed items and decrease for the negatively framed items. The mean ability estimates of each category for the items on the two scales were examined and the items with reversed average WLE are provided in Table 2.

Item P4 was discussed in the previous section. The statement of item S17, "I am told no matter how hard I try, I can't do well in math," is one of the most difficult items for students to endorse on the social persuasion scale. Furthermore, it has the most difficult threshold 3 among all threshold 3s for students to indicate "strongly agree." The discrimination index of this item is also very low (.05).

Content Validity. A concordance between the theoretical expectations of the item difficulty and the actual item estimates provides evidence for content validity (Cronbach and Meehl, 1955). It is not uncommon for students to overestimate their ability. Therefore, items such as, "Whatever I do, I didn't do as well as others" and "I try very hard, but still don't do well in school" were expected to have low endorsement rates, while items such as "I get high scores in most exams" and "Schoolwork is not so hard for me" were expected to be easy for students to endorse.

Both previous research (Mueller and Dweck, 1998) and daily life experiences suggested that many children are frequently praised and encouraged by their parents and teachers. For the social persuasion scale, students were expected to highly endorse items such as "If I fail at something, my parents encourage me to try again" and "I am often told that as long as I work hard enough, I

will be successful." Conversely, items such as "I am told no matter how hard I try, I can't do well in math" and "My parents seldom praise me for my schoolwork" were expected to have a very low number of endorsements from the students. Items describing positive persuasion were expected to be endorsed more frequently than items describing negative persuasion.

Spearman's rank order correlation provides information on the consistency between the expected item difficulty rank and the empirical one:

$$\text{Spearman's } \rho = 1 - \frac{(6 \sum D^2)}{I * (I^2 - 1)}, \quad (3)$$

with D being the rank difference and I being the number of items on each scale, which is 13 for the past performance scale and 17 for the social persuasion scale. The Spearman coefficient is .80 for the past performance scale and .75 for the social persuasion scale, which suggests that the empirical evidence well supports the theoretical item ranks.

External Validity. The correlation between the two scales is .78, which is reasonably high given that students who have had more successful past experience are more likely to receive positive feedback. When the instrument is expected to predict other external variables, a strong relation between this instrument and these external variables can be used as validity evidence (Messick, 1995). According to previous research (Ozyurek, 2005; Richard, 2005; Weinberg, et al., 1992), past performance and social persuasion are important predictors for self-efficacy belief. Therefore, the relationship between these two source scales and the general self-efficacy belief scale described above should provide some external validity

Table 2
Items with disordered average WLE

Scale	Item No.	Average Weighted Likelihood Estimates (Number of Students)			
		Score 0	Score 1	Score 2	Score 3
Past Performance	P4	-.27 (74)	-.30 (50)	-.20 (34)	.28 (5)
Social Persuasion	S17	-.06 (69)	-.37 (28)	-.21 (54)	.49 (4)

evidence for these two scales. The correlation between the past performance scale and the general self-efficacy belief scale is .43, and .51 between the social persuasion scale and the general self-efficacy belief scale. According to Schumacker (1996), measurement error lowers the correlation coefficient below the level it would have reached if the measures were precise. Therefore, the estimation of correlations should factor in the measurement errors and this process is referred to as disattenuation. This formula:

$$R_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} * r_{yy}}}$$

was used to estimate the error-controlled correlation coefficient, where r_{xy} is the unadjusted correlation coefficient, r_{xx} and r_{yy} are the reliability coefficient for measure X and measure Y , respectively. After the disattenuation, the correlation becomes .53 for the past performance scale and .66 for the social persuasion scale. Therefore, these two scales have reasonably satisfactory external validity after controlling for measurement errors.

Reliability

Internal Consistency. Internal consistency investigates the overall degree to which a subset of items are consistent with other sets of items. The Cronbach's alpha for the past performance scale is .85 and .80 for the social persuasion scale. By generally accepted psychometric standards (e.g., Nunnally and Bernstein, 1994), .70 is the minimum acceptable level of internal consistency reliability for use of a scale in subsequent research. Clearly, the alpha coefficients of both instruments investigated in this study meet this criterion.

Person Separation Reliability. The person separation reliability is .93 for the past performance scale and .91 for the social persuasion scale. There is evidence that these two scales can differentiate well between students with successful or not so successful past performances, and students who have received positive or non-positive feedback.

Discussion

To quantify the sources of self-efficacy beliefs, we developed and validated two scales for the two sources of self-efficacy belief: past performance and social persuasion. Both scales focused on the academic aspect of past student mastery experiences and the feedback they received from different sources (i.e., teachers, parents, and peers).

The Rasch rating scale model was used to analyze the data. Multiple sources of validity and reliability evidence were gathered. Results indicate that these two scales demonstrated satisfactory psychometric properties. One item (Item P4) was removed from the past performance scale due to large misfit, reversed average WLE, and low discrimination index (.07). The finalized version consists of 12 items for the past performance scale and 17 items for the social persuasion scale.

The development of these two scales allows the quantification of the kind of past performance and social persuasion students have experienced to be expressed. The enhanced knowledge about student past academic achievement allows teachers to identify the level of efficacy beliefs of their students. Efficacy belief has been demonstrated to be an important factor in predicting student academic performance in many previous studies (Jinks and Morgan, 1999; Kim and Lorschach, 2005; Pajares and Johnson, 1996; Schunk, 1991; Smith, 2001). Thus, it is essential for students to maintain a positive level of efficacy belief, as this makes them less likely to give up when encountering difficulties. These two scales can be used to track the possible reasons for low efficacy levels for some students. Information gathered from these two scales can also facilitate teachers and school counselors in decisions about whether a student needs extra help to strengthen their efficacy belief. For instance, if counselors decide that unsuccessful past performance is the major factor affecting academic self-efficacy beliefs, they may develop counseling strategies that provide opportunities for students to build successful master-experience (Bandura, 1986). Similarly, if responses from the students indicate that they did not receive enough positive persuasion, teachers

and parents should be made aware of this potential problems and should think of ways to be more positive and encouraging. Before a robust sense of efficacy is established, it is necessary that a positive learning environment is created to help students maintain and strengthen their academic efficacy beliefs.

As promising as these two scales appear, they have some potential limitations. First, due to the specific nature of the sample, the results may not be able to generalize to a larger population. In this study, the pilot data were collected from two public schools in northern California. The findings may not be universally true when the past performance and social persuasion scale are applied to other public schools in California, private schools in California, or schools out of California. When applying these two instruments, users should be aware of this caveat. When using these instruments to evaluate the sources of student efficacy beliefs, the proximity between the users' sample and the pilot sample examined in this study should be carefully considered. To further validate these instruments, a more heterogeneous sample consisting of geographically-diverse participants is needed. This way, the sample represents the population that the items intended to measure, so that the instruments can be used in a broader context.

Second, student self-report was used for both these two instruments and the criterion measure. Although it was made very clear in the data collection procedure that the responses were anonymous and students were encouraged to provide the most accurate answers, there was no way to get around the issue of subjectivity. The criterion students used to evaluate themselves could vary among individuals. For example, one student may choose "strongly agree" to a statement while another chooses "agree" even if they are at the same level of the trait measured by that particular item. The same limitation applies to the criterion variable. For future research, it is important to gather information about student past performance and the kind of social persuasion they received from more objective sources such as parents, teachers,

or peers. In addition, to increase the objectivity of the criterion measure, teacher and parent ratings of student self-efficacy, and official school documentations (e.g., GPA) can be used to cross-validate students' self report.

Having said that, as a pilot effort, this study contributes to the understanding of how academic efficacy beliefs were developed through previous mastery experience and social persuasions among middle school students. The instruments developed in this study can be of use for school administrators and teachers in determining the sources of students' efficacy level, as they make efforts to boost students' efficacy beliefs.

Acknowledgements

The authors gratefully acknowledge help from Professors Wen-Chung Wang and Martin Covington, for their constructive comments and suggestions on an earlier draft of this paper.

References

- Adams, R. J., and Khoo, S. T. (1996). *Quest*. Melbourne, VIC, Australia: ACER.
- Anderson, S. L., and Betz, N. E. (2001). Sources of social self-efficacy expectations: Their measurement and relation to career development. *Journal of Vocational Behavior*, 58, 98-117.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- 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., Claudio, B., Gian, V. C., and Conchetta, P. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72, 187-206.
- Betz, N. E., and Klein, K. (1996). Relationships among measures of career self-efficacy, generalized self-efficacy, and global self-esteem. *Journal of Career Assessment*, 4, 285-298.

- Chase, M. A. (1998). Sources of self-efficacy in physical education and sport. *Journal of Teaching in Physical Education, 18*, 76-89.
- Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Donnay, D., and Borgen, F. (1999). The incremental validity of vocational self-efficacy: An examination of interest, self-efficacy, and occupation. *Journal of Counseling Psychology, 46*, 32-47.
- Embretson, S., and Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*, 343-368.
- Feltz, D. L. (1988). Gender differences in the causal elements of self-efficacy on high avoidance motor tasks. *Journal of Sport Psychology, 10*, 151-166.
- Feltz, D. L. (1994). Self-confidence and performance. In D. Druckmand and R. A. Bjork (Eds.), *Learning, remembering, believing* (pp. 173-206). Washington, DC: National Academy Press.
- Feltz, D. L., and Riessinger, C. A. (1990). Effects on in-vivo emotive imagery and performance feedback on self-efficacy and muscular endurance. *Journal of Sport and Exercise Psychology, 12*, 132-143.
- Jinks, J., and Morgan, V. (1999). Children's perceived academic self-efficacy: An inventory scale. *Clearing House, 72*, 224-230.
- Kim, J., and Lorsbach, A. W. (2005). Writing self-efficacy in young children: Issues for the early grades environment. *Learning Environments Research, 8*, 157-175.
- Lent, R. W., Lopez, F. G., and Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology, 38*, 424-430.
- Lucas, J. L., Wanberg, C. R., and Zytowski, D. G. (1997). Development of a career task self-efficacy scale: The Kuder task self-efficacy scale. *Journal of Vocational Behavior, 50*, 432-59.
- McAuley, E. (1985). Success and causality in sport: The influence of perception. *Journal of Sport Psychology, 7*, 13-22.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp.13-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Mueller, C. M., and Dweck, C. S. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology, 75*, 33-52.
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Organization for Economic Co-operation and Development (2005). *PISA 2003 technical report*.
- Ozyurek, R. (2005). Informative sources of math-related self-efficacy expectations and their relationship with math-related self-efficacy, interest, and preference. *International Journal of Psychology, 40*, 145-156.
- Pajares, F., and Johnson, M. J. (1996). Self-efficacy beliefs and the writing performance of entering high school students. *Psychology in the Schools, 33*, 163-175.
- Pajares, F., and Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 26*, 426-443.
- Pajares, F., and Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving. *Journal of Educational Psychology, 86*, 193-203.
- Pajares, F., and Miller, M. D. (1995). Mathematics self-efficacy and mathematics outcomes: The need for specificity of assessment. *Journal of Counseling Psychology, 42*, 190-198.
- Paulsen, A. M., and Betz, N. E. (2004). Basic confidence predictors of career decision-making self-efficacy. *Career Development Quarterly, 52*, 354-362.
- Ramalingam, V., and Wiedenbeck, S. (1998). Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *Journal of Educational Computing Research, 19*, 367-381.
- Richard, J. D. (2005). An empirical investigation of sources of application-specific computer-self-efficacy and mediators of the efficacy-performance relationship. *International Journal of Human-Computer Studies, 62*, 737-758.
- Schumacker, R. E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions, 10*, 479.
- Schunk, D. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207-231.
- Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist, 25*, 71-86.
- Schunk, D. H., and Rice, J. M. (1993). Strategy fading and progress feedback: Effects on self-efficacy and comprehension among students receiving remedial reading instruction. *Journal of Special Education, 27*, 257-276.
- Sherer, M., and Maddux, J. E. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51*, 663-671.
- Shell, D. F., Colvin, C., and Bruning, R. H. (1995). Self-efficacy, attributions, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology, 87*, 386-398.
- Smith, S. M. (2001). The four sources of influence on computer self-efficacy. *Delta Pi Epsilon Journal, 43*, 27-39.
- Smith, E. V., Wakely, M. B., Kruif, R. L., and Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement, 63*, 369-391.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427-45.
- Weinberg, R. S., Grove, R., and Jackson, A. (1992). Strategies for building self-efficacy in tennis players: A comparative analysis of Australian and American coaches. *The Sport Psychologist, 6*, 3-13.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wise, J. B., and Trunnell, E.P. (2001). The influence of sources of self-efficacy upon efficacy strength. *Journal of Sport and Exercise Psychology, 23*, 268-280.
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transaction, 8*, 370.
- Wright, B. D., and Masters, G. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Wright, B. D., and Stone, M. H. (1988). *Validity in Rasch measurement*. Research memorandum 54. Chicago: University of Chicago, Department of Education, MESA Statistical Laboratory.
- Wu, M. L., Adams, R. J., and Wilson, M. R. (1998). *ConQuest*. Melbourne, VIC, Australia: ACER.

Appendix

SD: Strongly Disagree D: Disagree A: Agree SA: Strongly Agree

Past Performance Scale

P1	I have been doing very well in reading.	SD	D	A	SA
P2	I get high scores in most exams.	SD	D	A	SA
P3	I often fail in exams.	SD	D	A	SA
P4	I seldom know the answers to the questions asked in class.	SD	D	A	SA
P5	My teachers really like my work.	SD	D	A	SA
P6	I learn all the subjects very well.	SD	D	A	SA
P7	Schoolwork is not so hard for me.	SD	D	A	SA
P8	I have always been a top student in class.	SD	D	A	SA
P9	I try very hard, but still don't do well in school.	SD	D	A	SA
P10	I enjoy doing my schoolwork.	SD	D	A	SA
P11	I spend a lot of time on schoolwork, but still don't do well.	SD	D	A	SA
P12	I have been doing well in math.	SD	D	A	SA
P13	Whatever I do, I didn't do as well as others.	SD	D	A	SA

Social Persuasion Scale

S1	If I don't do well in exams, my parents encourage me to try again.	SD	D	A	SA
S2	My parents always say I am not a smart kid.	SD	D	A	SA
S3	My parents never yell at me when I fail in exams.	SD	D	A	SA
S4	My teachers think highly of my learning ability.	SD	D	A	SA
S5	My teachers tell me I have low math ability.	SD	D	A	SA
S6	My teachers say I am slow in learning new subjects.	SD	D	A	SA
S7	My teachers always say I do a good job in class.	SD	D	A	SA
S8	I am punished by my parents when I don't do well in school.	SD	D	A	SA
S9	My parents seldom praise me for my schoolwork.	SD	D	A	SA
S10	I have been often told that I am really smart.	SD	D	A	SA
S11	My friends say I do better than them in schoolwork.	SD	D	A	SA
S12	I am told I learn things fast.	SD	D	A	SA
S13	I am often told that as long as I work hard enough, I will be successful.	SD	D	A	SA
S14	Other people think I can easily learn new things.	SD	D	A	SA
S15	Other people think I am not as clever as other students.	SD	D	A	SA
S16	I am told I have a gift for math.	SD	D	A	SA
S17	I am told no matter how hard I try, I can't do well in math.	SD	D	A	SA

*(Appendix continued on the following page.)**(Appendix continued from the previous page.)**General Self-efficacy Scale*

G1	When I make plans, I am certain I can make them work.	SD	D	A	SA
G2	I avoid facing difficulties.	SD	D	A	SA
G3	I give up easily.	SD	D	A	SA
G4	When I decide to do something, I go right to work on it.	SD	D	A	SA
G5	When trying to learn something new, I soon give up if I am not successful from the beginning.	SD	D	A	SA
G6	If something looks too hard, I will not even bother to try it.	SD	D	A	SA
G7	One of my problems is that I cannot get down to work when I should.	SD	D	A	SA
G8	When I set important goals for myself, I almost never achieve them.	SD	D	A	SA
G9	I give up on things before completing them.	SD	D	A	SA
G10	Failure just makes me try harder.	SD	D	A	SA
G11	When unexpected problems come up, I don't handle them well.	SD	D	A	SA
G12	If I can't do a job the first time, I keep trying until I can.	SD	D	A	SA
G13	I do things on my own without help.	SD	D	A	SA
G14	I feel not so confident about my ability to do things.	SD	D	A	SA
G15	I avoid trying to learn new things when they look too difficult for me.	SD	D	A	SA
G16	When I have something unpleasant to do, I stick to it until I finish it.	SD	D	A	SA
G17	I do not seem capable of dealing with most problems that come up in my life.	SD	D	A	SA

*Note: This item was removed for showing large misfit, reversed average WLEs, and low discrimination.