ASSESSING SCIENTIFIC AND TECHNOLOGICAL SELF-EFFICACY: A MEASUREMENT PILOT

Diane Jass Ketelhut, Harvard University Graduate School of Education
Diane@alumni.brown.edu

Abstract

The purpose of this study was to develop a specific measure of self-efficacy in science inquiry and technology for middle school students. The initial survey of 61 items was piloted online first with a sample of 100 students. Revisions were made based on reliability and validity evidence, using Cronbach's alpha, Principal Component Analysis and correlational evidence. Two subscales of the survey were then implemented with over 2000 students participating in a technology in science research project. The final version of the survey consists of six independent sections on self-efficacy in science inquiry and technology (science inquiry, using the Internet to find information, general computer usage, synchronous chat use, videogaming and computer gaming).

Problem

Bandura (1977) coined the term "academic self-efficacy" for the belief that one can succeed in learning a particular type of content. Unfortunately, many students lose self-efficacy in science as they go through school, as indicated by the *National Assessment of Educational Progress* (National Center for Educational Statistics, 2000). An item-stem, measuring self-efficacy, in this large survey asserts "all can do well in science if they try." 82% of fourth-graders agreed with this statement; however, only 64% of eighth-graders and a mere 44% of twelfth-graders agreed (NCES, 2000).

This paper discusses the development, piloting and revision of a new instrument for measuring the academic self-efficacy of students who study scientific inquiry in a technology-intensive learning experience. My initial motivation for the creation of this new instrument was the need to measure self-efficacy in an NSF-funded project designed to investigate the motivational effects of a multi-user virtual environment (MUVE) on the science achievement of middle-school students. However, I believe that individual sections of the new instrument can contribute in general to the more-finely discriminated measurement of self-efficacy. The next section reviews the background and context for this new instrument. Then, the research design for the development of the pilot is described outlining sample, measures, and procedures. This is followed by the data analysis, and results from the larger sample in which it was further tested.

Theoretical Underpinnings

Bandura (1977) defined self-efficacy as the belief that one can successfully perform certain behaviors. He hypothesized that individuals obtain information about their self-efficacy in four ways. First, students' own performances affect their self-efficacy: students who successfully graph data will feel more confident when again asked to graph data. Secondly, students' vicarious experiences affect their self-efficacy: when a student sees a peer successfully graphing data, she may feel more confident when asked to graph data herself. Thirdly, students' self-efficacy can be affected by others' verbal persuasion: a teacher may persuade a student that she can successfully graph data, and thus she approaches the next graphing task confidently. The fourth factor is emotional arousal: for example, a student's confidence in approaching a graphing exercise inversely depends on his level of anxiety induced by that assignment. All of these experiences can affect self-efficacy either positively or negatively.

Since self-efficacy was defined in the late 1970s, many researchers have designed instruments to measure the construct. Some of these instruments are framed generally while others are designed for administration in specific settings and intellectual contexts. Midgeley et al (2000) developed a general measure of self-efficacy for administration to K-12 students. Within the instrument, item-stems are written generically: “I am certain I can master the skills taught in class this year.” Miltiadou & Yu (2000) created an online instrument, designed specifically to measure the self-efficacy of college students using online communications. Here, item-stems were specific: “I would feel confident reading messages from one or more members of the synchronous chat system.”

Bandura (1986, 1997) and others (Miltiadou & Yu, 2000; Pajares, 1996) have argued that it is not sufficient to measure self-efficacy globally; this construct must be measured in a context-specific way. Therefore, Bandura suggests that measures of self-efficacy be designed as specific to the task being performed as possible. While some researchers have cast doubt on this (Bong, 1996), the strength of the relationship between self-efficacy and performance appears to weaken when more global instruments are used to measure it (Pajares, 1995).

There are currently no context-specific instruments available for measuring self-efficacy for middle-school students using an inquiry process to learn aspects of experimental scientific design, either with or without the associated use of learning technologies. A more specific instrument, a *Self-efficacy in Technology and Science* (*SETS*) instrument, is needed to measure academic self-efficacy reliably and validly in research on scientific inquiry, including within a technology-based learning experience. This investigator’s work deals with learning higher order scientific inquiry skills via a Multi-User Virtual Environment (MUVE), which has implications for the larger context of game-related or simulation-based science curricula. Such an instrument could also find use in the larger context of game-related or simulation-based science curriculum, as well as technology-based learning experiences in general.

Although some researchers may use unpiloted survey instruments in their research, conducting a measurement pilot to establish reliability and validity of a new instrument in the intended empirical context is considered crucial by most (Litwin, 2003; Popham, 1981). For example, in creating their online self-efficacy instrument for college students, Miltiadou & Yu (2000) estimated internal consistency reliability, as measured by Cronbach’s alpha, to establish the precision of their instrument. Midgeley et al (2000) also estimated Cronbach’s alpha reliability for their instrument.

In addition, self-efficacy researchers have adopted a variety of methods for demonstrating the validity of their instruments. Midgley et al (1998) provided evidence of their instrument’s construct validity by examining correlations between scores on their instrument and scores on instruments measuring other closely-related constructs, as indicated in the literature. To create a measure of general self-efficacy Jinks and Morgan (1999) demonstrated the convergent validity of their instrument by correlating self-efficacy scores with students’ self-reported grades. Finally, they demonstrated content validity of their scale by having experts review and comment on their instrument.

Design and Procedure

Research questions.

In the development of this instrument, the following research questions were addressed:

RQ1. What is the internal consistency reliability of the *SETS* instrument?

a. What is the internal consistency reliability of each subsection of *SETS*?
b. What contribution does each item on *SETS* make to the overall and subscale internal consistency reliability?

RQ2. Does the *SETS* instrument demonstrate:

a. Content validity?
b. Construct validity?

Pilot Sample.

SETS was first administered to a convenience sample of 98 fifth through eighth grade students. In order to identify sample children for the measurement pilot, volunteer teachers were solicited who would permit me to administer the instrument to their students in the context of their science classes. The final instrument was then administered to over 2000 middle-school students participating in a research project on learning experimental design using virtual environments.

Measures.

The draft instrument measured self-efficacy in five sub-contexts: (a) videogame-playing, (b) synchronous chat use, (c) general computer use, (d) inquiry science processes, and (e) generic science learning. These subsections represent content and process areas utilized in science curricula employing emerging technologies. Each of the first four sections started with 15-16 item-stems and Likert-type response statements. The generic science section consists of six items from the already-validated *Patterns of Adaptive Learning Scales* (Midgley et al, 2000) and was included for comparison purposes. The draft instrument underwent evaluation by a team of experts that are members of a larger NSF-funded technology project's consultant panel. Each expert evaluated the instrument for content, clarity and appropriateness. Item-stems were modified based on their feedback.

Procedures.

After the experts evaluated *SETS*, the sequence of 67 statements in the five subsections (plus ten demographic questions) were randomized and administered to sample children online, supervised by the participating teachers, using *Survey Monkey* (http://www.surveymonkey.com/). Students finished the survey in a single class period.

Findings

RQ1. What is the internal consistency reliability of the SETS instrument?

I conducted both an item analysis and a principal components analysis to estimate the internal consistency reliability of the *SETS* scale, including the reliability of its subscales, and to estimate the contribution that each item made to each subscale. Item-stems identified by those tests as measuring a different construct were removed and examined for content differences; the internal consistency reliability was recalculated for the remaining items.

For the sections on self-efficacy in inquiry science and synchronous chat use, the results of the internal consistency reliability supported removal of 3 and 5 items, respectively. The self-efficacy in inquiry science ended with 12 statements and a reliability of .90. The analysis of the self-efficacy in synchronous chat use section resulted in 10 statements with a reliability of .92.

The section on videogame-playing initially included statements about videogames and computer games. Upon evaluation, the results from the pilot implementation indicated that these were separate constructs; the items were therefore analyzed separately. At the end of the analysis, the self-efficacy in videogaming section consisted of 8 statements with a reliability of .93; the self-efficacy in computer gaming section consisted of 5 statements with a reliability of .85.

A similar situation was discovered with the section on self-efficacy in computer usage. Results of the PCA indicated that this section was measuring more than one construct. Examining the statements separated by the results of Cronbach's alpha and PCA indicated that the statements belonged in two groups. The first group consisted of 11 statements, measuring self-efficacy in general computer usage with a reliability of .81. The second group consisted of 5 statements, measuring self-efficacy in using the Internet to find information. This group had a reliability of .79.

Based on the results of the internal consistency reliability, the final version of SETS consists of six new self-efficacy constructs, not the original four. Furthermore, the self-efficacy in inquiry science section was re-administered to 1100 students; results confirm the above analysis, showing an internal consistency reliability of .86.

RQ2. Does the SETS instrument demonstrate content validity and construct validity?

To establish content validity, the final version of SETS was shown again to the content experts. While two minor concerns were raised, overall they concurred that the statements were accurate measures of the indicated content.

To establish construct validity, correlations among each student's score on the individual sections and various collected demographic information were calculated. These correlations were compared with what was expected based on the literature. While the validity process is an ongoing one, initial indications are that these sections do indeed measure what they are intended to measure. For example, the correlation between self-efficacy in videogaming and gender is high ($r = .59$), the direction indicating that males having a significantly higher self-efficacy than females ($p<.0001$), supporting indications from the literature.

Conclusion and Importance of study

Bruce M. Alberts, chair of the National Research Council, stated in 1995, "We've managed to turn people off of science by making it some kind of rote learning exercise" ("Panel Urges Shift of Focus for School Science Courses," 1995). With the rapidity of new discoveries in science and technology, it is imperative that high school graduates be able to evaluate, not blindly accept, information offered in newspapers, television magazines and by politicians. While not all students need to become scientists, they all need a measure of scientific literacy. We cannot allow our students to graduate from high school feeling as if science is beyond their capabilities. "We must give a high priority to making scientific and technical topics accessible and understandable, as well as fascinating and exciting, to the general public" (Massey, 1988, p 18).

A new approach is needed to increase students' interest and belief in their own capabilities in science. Can academic self-efficacy be affected by technology? If so, how and for whom? Once students perceive themselves as more competent, they can be challenged with more complex material as their effort and perseverance will increase (Bandura, 1977), making our students more competitive globally. Determination of a valid measure of self-efficacy is the first step in that research. This instrument contributes to the general self-efficacy research community, as well as to the science and technology research community.

References

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1986). *Social Foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bong, M. (1996). *Perceived Similarity among Tasks and Generalizability of Academic Self-Efficacy* (REPORTS - Research/Technical. SPEECHES, CONFERENCE PAPERS.).

Jinks, J., & Morgan, V. (1999). Children's Perceived Academic self-efficacy: An Inventory Scale. *The Clearning House, 72*(4), 224-230.

Litwin, M. S. (2003). *How to Assess and Interpret Survey Psychometrics* (2nd ed. Vol. 8). Thousand Oaks: Sage Publications.

Massey, W. E. (1988). Making Science Accessible. *Liberal Education, 74*(2), 16-19.

Midgley, C., Kaplan, A., Middleton, M. Maehr, M., Urdan, T., Anderman, L., Anderman, E., Roeser, R. (1998). The Development and Validation of Scales Assessing Students' Achievement Goal Orientations. *Contemporary Educational Psychology, 23,* 113-131.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, & M., K., A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS).* Ann Arbor, MI: U. of Michigan.

Miltiadou, M., & Yu, C. H. (2000). *Validation of the Online Technologies Self-Efficacy Scale (OTSES)*.

National Center for Educational Statistics.*The Nation's Report Card*. Available: http://nces.ed.gov/nationsreportcard/ [2003, October 12].

Pajares, F. (1995). *Self-Efficacy in Academic Settings.* Paper presented at the American Educational Research Association, San Francisco, Ca.

Pajares, F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*, [Internet]. Available: http://www.emory.edu/EDUCATION/mfp/aera2.html [2004, February 21].

Panel Urges Shift of Focus for School Science Courses. (1995, December 7). *New York Times*.

Popham, W. J. (1981). *Modern Educational Measurement*. Englewood Cliffs, NJ: Prentice-Hall, Inc.