Measuring Change In Efficacy

Everett V. Smith, Jr.
University of Illinois at Chicago
Kimberly A. Lawless
Utah State University
Leslie Curda and Steven Curda
University of Toledo

"I think I can, I think I can"

(Piper, W. (1954). The Little Engine That Could).

As we all remember, the little engine succeeded in climbing the hill. Not because it knew it could succeed based on past performance, but because it believed it could succeed. This self-belief in one's ability to perform a specific task is known as self-efficacy (Bandura, 1986). It is a person's "I can" or "I cannot do" belief. It is not concerned with the skills one has, but with the judgments of what one can do with the skills one possesses.

Why Measure Self-efficacy?

A primary purpose of training programs is to convey a body of knowledge which can then be applied in the future. How effective is instruction in conveying this body of knowledge? The typical evaluation of instruction only assesses acquisition of knowledge while overlooking self-efficacy. Self-efficacy influences persistence and motivation, important outcomes to any training program. In conjunction with achievement data, self-efficacy measures can serve as an important part of program planning and evaluation (Owen, 1991), (indicating areas where individuals do not possess sufficient confidence in their ability to perform specific tasks, either prior to, during, or following a program of instruction.) Perceived weaknesses can suggest more efficient instruction. As Lusardi and Smith (1997) discuss, self-efficacy measures can be better indicators of use of recently acquired knowledge than outcome measures. Evidence that knowledge has been learned (outcome measures) is not evidence that knowledge will be applied. A training program is limited if alteration of behavior is achieved but the learner is not endowed with confidence to engage in the behavior at some future point in time.

Background

We evaluated change in self-efficacy for a group of undergraduate preservice teachers (n=48) enrolled in an instructional technology course. One of the issues facing the educational field today is the incorporation of new teacher competencies into existing requirements for teacher certification (e.g., ISTE, NCATE). Many of these competencies deal with teachers' ability to utilize a variety of technologies. As these new competencies become more common and teacher preparation programs become accountable for graduating students with these skills, courses must be developed and evaluations conducted to ensure that competencies are being met. As these courses are developed, teacher educators must deal with issues that may interfere with their students' willingness to engage in new experiences with technology. Preservice teachers' efficacy in the use of technologies is a key factor that will influence whether they are willing to integrate technology into the curriculum.

The Classroom Technology Questionnaire (CTQ) was designed to assess students' self-efficacy in implementing instructional technologies. The CTQ consists of 14 items, each focusing on a different form of instructional media. For each type of media, a definition was provided to help all respondents to respond from a common frame of reference. For each form of media and item stem, respondents were asked the following question: "Imagine you are teaching RIGHT NOW. How skillful do you feel about using this type of media in your classroom RIGHT NOW?" Responses were collected on a 7-point Likert type scale, with poles labeled NOT AT ALL SKILLFUL (1) and EXTREMELY SKILLFUL (7).

The evaluation

Program evaluations typically use aggregate data and assess the impact of a program using a sample dependent t-test. This results in an evaluation of whether the group mean is significantly different from pretest to posttest. For diagnostic pur-

poses, there is a need to be able to locate individuals who are different from the group both prior to instruction and upon completion of the course. Further, for valid pre-post interpretations, the potential problem associated with the functioning of the items, which may be interpreted differently at each measurement occasion by the respondents, needs to be investigated.

Results

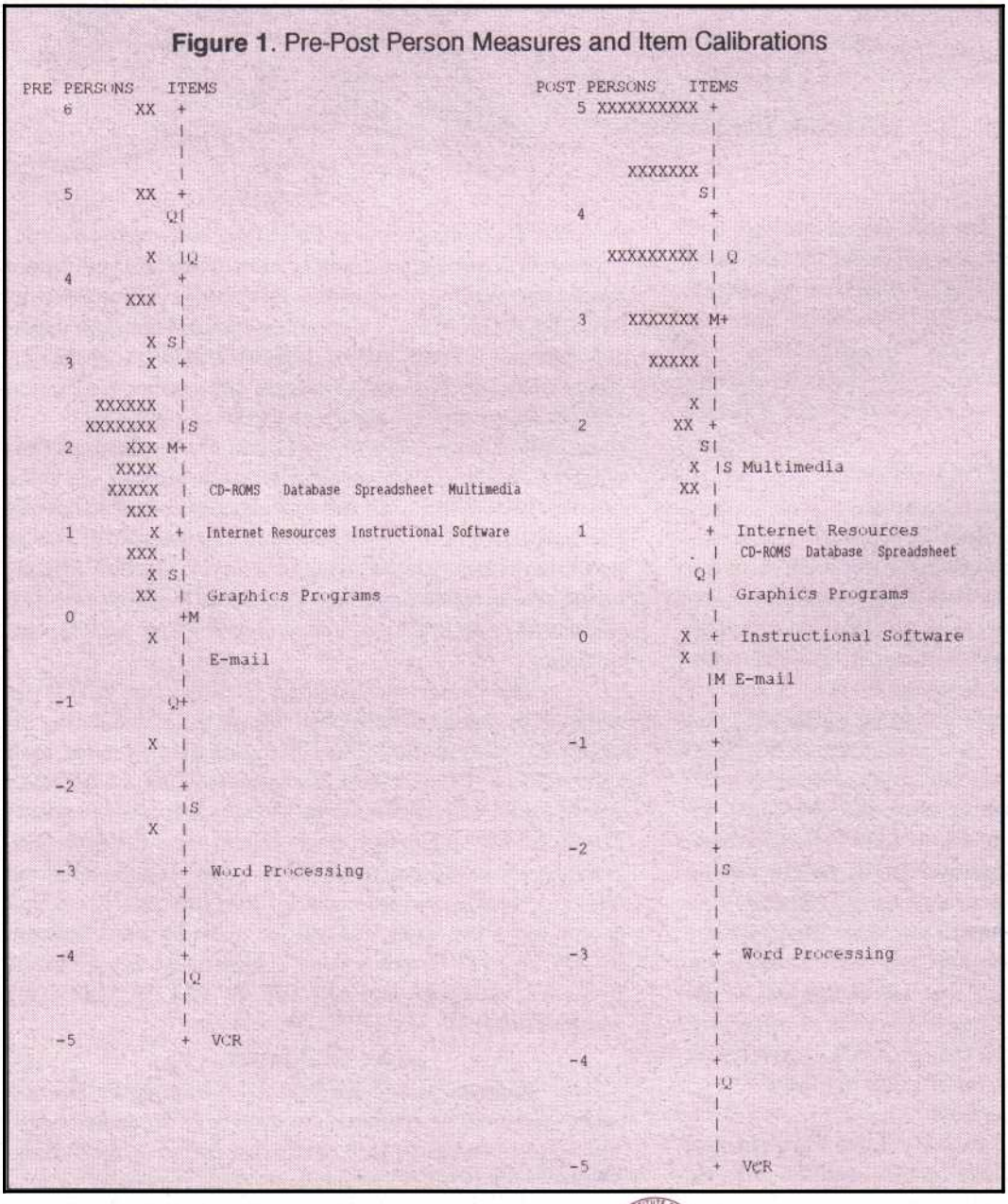
The results presented are based on ten, of the fourteen, items that were found to fit the Rasch Rating Scale Model. It was also found that a 4-point scale better represents the data than the original 7-point scale.

The traditional method of assessing change using a dependent t-test demonstrated statistically significant gains for the group from pretest to posttest (**Table 1**). This method of program evaluation has two limitations. First, changes in the underlying variable are not investigated. If the variable being measured is not the same from pretest to posttest, evaluation of change is meaningless (see Wright, 1996). Second, rather than concentrating on group differences, it would be of greater value to see which individuals demonstrated statistically significant gains or losses.

Table 1
Results of pre-post analysis at the group level

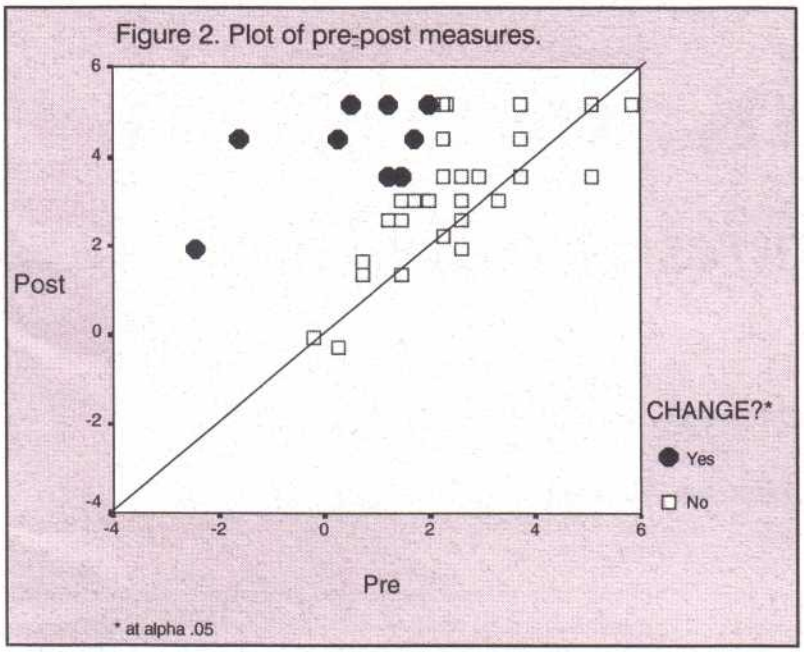
	Mean	SD	Mean Difference	SD	t	p
Pre measure	2.08	1.67	1.36	1.66	5.29	.001
Post measure	3.44	1.48				

Figure 1. Pre-Post Person Measures and Item Calibrations



Fortunately, Rasch measurement can be used to address both of these deficiencies. **Figure 1** compares the variable being measured at pretest and posttest. Notice that several of the items maintain their location on the variable, indicating stability (invariance) of the item calibrations. This type of evidence is required in order to make valid pre-post comparisons. **Figure 1** also displays the shift in person measures (the shift in the group mean labeled 'M'), and pictorially represents the results of the dependent t-test (Note **Figure 1** represents calibration of all available data from pretest (n=48) and posttest (n=46) administrations, while the dependent t-test is based on complete data only (n=42)).

Rasch measurement also produces standard errors for



each measure. This distinctive advantage over Classical Test Theory allows for the statistical comparisons of pre-post scores at the individual rather than group level. Using this information, one is able to target individuals that did not display statistically significant gains in self-efficacy and those that demonstrated reductions for further self-efficacy enhancing activities. **Figure 2** demonstrates analysis of change at the individual level. Darkened squares above the identity line indicate statistically significant gains for those individuals from pretest to posttest. Nine students demonstrated statistically significant gains, none statistically significant reductions (fortunately). This information could prove of great value in evaluating a current course or planning a future course, if follow-up procedures are undertaken to investigate how and why the program benefited several individuals while seemingly not affecting others.

References

Bandura, A. (1986). Social Foundations of Thought and Actions: A Social Cognitive Theory. New Jersey: Prentice-Hall, Inc.

Lusardi, M.M., & Smith, Jr., E.V. (1997). Development of a scale to assess concern about falling and applications to treatment programs. Journal of Outcome Measurement, 1, 34-55.

Owen, S.V. (1991, April). Using self-efficacy in program evaluation. Paper presented at The Training Excellence: A Conference on training practitioners in the drugs/alcohol field, Canterbury, England.

Piper, W. (1954). The Little Engine that Could. New York: Platt and Munk.

Wright, B.D. (1996). Time 1 to Time 2 Comparisons. Rasch Measurement Transactions, 10(1), 478-479.

Kimberly A. Lawless, Assistant Professor of Instructional Technology, Utah State University. Teaching areas include technology in education, learning theory, research methods, and product evaluation. Research interests include reading comprehension, hypertext processing, teacher beliefs, and instructional design practice. Hobbies include hiking, refinishing furniture, and spending time with family and friends.

L. K. Curda is an Assistant Professor and Program Coordinator of Educational Technology at the University of Toledo. Her research activities include collaborating with university and public school colleagues to investigate groups of ecological variables to explicate the influences they have on teachers' expectancies and goals and their performance in the classroom related to technology integration in content areas. Much of her early research has been focused on developing valid and reliable measures for investigating variables associated with change management and diffusion of innovations that can be used to evaluate the effectiveness of preparation programs in developing motivated teachers, and to profile a school's practicing teachers in order to develop and implement needed interventions to increase teacher motivation for and implementation of technology-related initiatives. Email: lcurda@utoledo.edu

"When the Rasch model is intended to hold because of its special measurement properties, failure of the data to conform to the model implies further work on the substantive problem of scale construction, not on the identification of a more complex model that might account for the data."

David Andrich
in *Rasch Models for Measurement*.
1988. p.86. Newbury Park, CA: Sage.