Validity of Retrospective Pretest Methodology

Alexander Blount, EdD

Those reviewing evaluations of training programs have a right to be skeptical. Much evaluation is superficial (Bamberger, et al. 2004; Brooks & Gersh, 1998). The opinion of an individual about how well the training program is delivered and how much they like the various aspects of the program is not related in any reliable way to each individual's gaining knowledge or skills. Individuals' evaluations of their own skill development is also notably unreliable (Albanese, et al., 2006; Davis, et al., 2006). It is very hard for a person to reliably assess their own skills in a domain of expertise. Experimental and quasi-experimental designs, particularly those with control groups are difficult or impossible for most training settings to utilize (Brooks & Gersh, 1998).

One could reasonably wonder whether aggregating a group of participants' scores would make the situation worse. If a person's assessment of their own skill development is unreliably related to their observed skill demonstrations, wouldn't a group score introduce even more unreliability? While an individual's assessment of their skills may be unreliable, when aggregated, the sum of individuals' assessments has proved to be quite reliable for assessing the training of the group (Bernthal, 1995).

It might seem reasonable to look for changes in the outcome for patients who are served by a participant in a training program. A participant who takes a training program to learn how to better make “widgets” ought to be able to make better widgets or make more widgets in a given time period without losing quality after they take the training than they could before. In that case, assessing the number and quality of widgets produced would be an excellent way of assessing the training program. In domains in which the expertise delivered or enhanced by the training program involves improved skills in relating with patients and working as part of a team in a complex multi-determined setting, correlating training success with client change becomes impossible under most circumstances. Too many other factors are involved to make a judgment about the effectiveness or ineffectiveness of a training program based on clinical change of the clients as observed by some measure.

For example, the organization which has sent the largest number of staff to the Certificate Program in Primary Care Behavioral Health subsequently received a national award for quality and innovation in care. It now has been designated as an exemplar of integrated primary care by the Academy for the Integration of Behavioral Health and Primary Care (http://integrationacademy.ahrq.gov/). Unfortunately, it is not possible for our Center to claim more than an unspecified small role in the success of their program.
We are left with a two-step inference process: 1. Did the trainees enact the expertise in ways closer to some standard after the training program than before? and 2. Has the expertise as defined by the standard been correlated with positive change for patients in more controlled studies?

In the last few years the science of evaluation of training programs has been moved forward by studies that have begun to identify reliable ways of measuring the acquisition of skills of participants. A process called retrospective pretest assessment has been shown to be a reliable measure of the success of a training program and to correlate well with pre and post ratings of the performance of participants made by observing experts (Goedhart & Hoogstraten, 1992; Terborg, et al. 1980; Pratt, et al. 2000). Retrospective pretest involves asking participants to assess their skill levels in the areas addressed by the training programs. After the program is completed, participants rate their skills as they remember them before the training and compare them to their ratings of their skills after the training experience. This is more reliable than having people actually rate their expertise at the beginning of the program and then again at the end of the program. When people rate their expertise before they take the program, they tend to underestimate how much there is to know, and so to overestimate their own level of learning or skill. This shift is called the response shift bias (Howard & Dailey, 1979). Once they have taken the training course, they have a better understanding of what is involved in the domain of expertise and tend to produce a more “humble” assessment of their skills in relation to what there is to master, reducing their assessment of their own learning or skill at the inception of the course. When people rate their expertise at the end of the course, rating both what they could do at the beginning and comparing it to the end, we have a much more informed and valid assessment (Goedhart & Hoogstraten, 1992; Terborg, et al. 1980; Pratt, et al. 2000).

We still need to address the question of whether a valid reporting of the experience of participants of their learning in a training program correlates with a change in the skill they show with patients. D’Eon and colleagues (D’Eon, et al. 2008) conducted a study training medical students to give feedback to patients. They were videotaped giving the feedback to standardized patients (actors who play the same patient with the same medical profile for purposes of training the helping professional) before and again after the training course. The videotapes were reliably scored (different raters gave a given tape the same or a similar score using a standard scoring system). The raters were blinded to whether a tape was of an interview before or after the course. Medical students assessed their own skills using the retrospective pretest methodology. The training was two half-days long. The pre-post assessment of their skill development correlated very well with the pre and post assessment of change in demonstrated skill as observed by the trained raters assessing their videotaped performance. There was a slightly larger effect reported from the pre-post study than from the observational comparison of expertise, but the correlation supported an interpretation that the assessment of the change of the group as a whole done by the retrospective pretest was a valid proxy for observed
change in enacted expertise. The rating of skill acquisition by participants correlated extremely well with demonstrated skill acquisition.

Conclusion:

Retrospective pretest methodology is a valid way to measure the change in skills that occurs when participants take a training program. Participants’ rating of their change of skills correlates well with observations of their change of skills in practice as rated by experts in the field.

References:


