

Online Scorer Training

Increasingly, technology is being employed to improve the effectiveness and efficiency of delivery, scoring, and reporting of large-scale assessments. Justifiably, when changes are made to testing procedures, stakeholders in the testing process want verification that the fidelity of inferences that are made about test scores is not threatened.

A somewhat recent application of technology in testing is the use of *distributed online training and scoring* for scorers of constructed-response test items. Typically, scorers are trained at a regional center, having scorers arrive at a predetermined time, conducting a face-to-face presentation of the scoring rubric, and then asking scorers to complete any qualifying requirements prior to working on the scoring project. That process is somewhat inefficient for several reasons. First, because scorers and trainers need to travel to the training location, there are travel costs that either the scorer or the company sponsoring the training must cover. Second, because the training is done in a group setting, the speed at which training can be delivered is restricted by the speed of the slowest individuals in the training group. As a result, some scorers, who might be trained more quickly, spend more time in training than may be necessary. Third, the materials for the training are typically paper-based, leading to both expense and utilization of natural resources. Fourth, due to the social context of group training, additional discussions may take place that may slow down the training process (e.g., discussion of logistical issues).

Conversely, online training and scoring offers several potential opportunities to increase effectiveness and efficiency. For example, online distributed training and

scoring (i.e., distribution of student responses via a network to scorers at remote locations) significantly increases the flexibility of work hours. Scorers can perform the work from home at hours that are convenient for their schedules, making it possible to recruit scorers who may not otherwise be available to participate in a scoring project that is conducted at a regional center. In addition, because training is delivered individually at a pace controlled by the scorer, training may be completed more quickly by most scorers. Also, resources can be conserved because paper-based documentation may not be necessary in online scoring.

"Online training and scoring offers several potential opportunities to increase effectiveness and efficiency."

In addition, scores can be entered directly into a database, with no intermediate data entry step, allowing for the scores to be available more quickly than is possible with scanned forms. Distribution of student responses can be more easily randomized in an online scoring system—there is no need to distribute booklets of responses to randomly chosen pairs of scorers as is often the case in paper-based scoring. This randomization helps minimize the risk of scorer pairing effects influencing decisions that are made about individual examinees. Finally, it may be easier to implement scorer monitoring practices in an online scoring system. Each score that is assigned by a scorer can be monitored in real time, and corrective actions can be taken immediately, and those procedures can also be automated, redirecting scorers to remedial training or corrective feedback

without the need for intervention by scoring facilitators.

Fortunately, research has been conducted which suggests that several of the potential benefits associated with online training of scorers and scoring may be realized while maintaining the fidelity of the training and scoring process. For example, studies of online training of scorers for a second language testing program have revealed that, when raters are trained online, assigned ratings may exhibit modest increases in reliability and slightly lower levels of rater effects when compared to those trained in a face-to-face setting. In addition, those studies revealed no strong preference on the part of scorers for one training medium versus the other (Elder, Barkhuizen, Knoch, & von Randow, 2007; Knoch, Read, & von Randow, 2007).

Another study, which compared self-paced delivery of scorer training to face-to-face presentation, suggested that a self-paced approach can result in training that is just as effective as face-to-face training while significantly increasing the speed of training completion. In that study, scorers in the self-paced group spent about two-thirds as much time training as those in the face-to-face group (Liao & Campbell, 2002).

At Pearson, two studies were conducted to take a comprehensive look at three modes of scorer training and scoring:

Standup/Regional (i.e., face-to-face training followed by online scoring at a regional center), **Online/Regional** (i.e., online training conducted at a regional center, followed by online scoring at that center), and **Online/Distributed** (i.e., online training conducted at distributed sites, typically the scorer's home, followed by online scoring at those sites). In each study, 40 scorers who had not used the rubrics adopted for the study were allocated to the three training/scoring

contexts (i.e., 120 scorers in each study) so that the groups were comparable in terms of demographics, education, and professional experience. In one study, scorers were trained and scored written essays, and in the other study scorers were trained and scored responses to three reading prompts. After completing training, scorers underwent qualifying testing and then scored 400 student responses using the scoring rubric for which they were trained. Researchers collected information about the speed of training and scoring, scorer performance on qualifying sets, the reliability and validity of the assigned scores, and rater perceptions of the effectiveness of and their satisfaction with the training procedures and materials, personnel, and scoring process and materials.

The results were consistent across the two studies. Overwhelmingly, standup training took more time than did online training—between two and three times longer.

“Overwhelmingly, standup training took more time than did online training—between two and three times longer.”

For example, online training took about four hours in writing, while face-to-face training took twelve hours. Similarly, each average amount of training time per prompt in reading took about two hours, while the face-to-face training delivery took about five hours. Remarkably, there were few other differences between the groups. Scoring, which was conducted via an online interface in all three groups, was completed in a comparable amount of time

in both content areas. Raters achieved very high levels of agreement between the scores they assigned to qualifying set responses and those assigned by scoring leaders—almost 80% exact agreement—in all three training/scoring groups in both content areas. Similarly, the quality of the assigned ratings was also high and was comparable between the three training/scoring groups. It is noteworthy that both groups of scorers who were trained online exhibited slightly higher inter-rater agreement as well as agreement with scoring leaders, although the between-group differences were not statistically significant. Finally, researchers observed no noteworthy differences between the three training/scoring groups in either content area concerning rater perceptions of the effectiveness of and satisfaction with the training and scoring process.

According to the results of this study, online scorer training as it is implemented by the Performance Scoring Center at Pearson, matches, and possibly exceeds, face-to-face training in terms of effectiveness. That study revealed no large difference between scorers trained in online and face-to-face contexts in terms of performance on qualification sets and operational scoring. The small differences that were observed favored the online group. Similarly, measures of rater perception of the training and scoring process revealed no appreciable differences in preference for or satisfaction with online or face-to-face training contexts. However, consistent with prior research concerning self-paced scorer training, these two studies indicate that both distributed and regional online training is considerably more efficient than face-to-face training. Specifically, training that took place in a face-to-face regional context took two to three times longer to complete than did online training. Hence,

it seems that online training may be as effective as face-to-face training while affording the possibility of a faster turn around time due to reduced training time.

-- Edward W. Wolfe, Ph.D.

REFERENCES

- Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing, 24*(1), 37-64.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing, 12*(1), 26-43.
- Liao, P.-j., & Campbell, S. K. (2002). Comparison of two methods for teaching therapists to score the test of infant motor performance. *Pediatric Physical Therapy: The Official Publication Of The Section On Pediatrics Of The American Physical Therapy Association, 14*(4), 191-198.