CLEAR Exam Review is a journal, published twice a year, reviewing issues affecting testing and credentialing. CER is published by the Council on Licensure, Enforcement and Regulation, 403 Marquis Ave., Suite 100, Lexington, KY 40502.

Editing and composition of this journal have been underwritten by Thomson Prometric, which specializes in the design, development, and full-service operation of high-quality licensing, certification and other adult examination programs.

Subscriptions to CER are sent free of charge to all CLEAR members and are available for $30 per year to others. Contact Stephanie Thompson at (859) 269-1802, or at her e-mail address, sthompson@mis.net, for membership and subscription information.

Advertisements and Classifieds (e.g., position vacancies) for CER may be reserved by contacting Stephanie Thompson at the address or phone number noted above. Ads are limited in size to 1/4 or 1/2 page, and cost $100 or $200, respectively, per issue.

Coeditor
Michael Rosenfeld, Ph.D.
Educational Testing Service
Princeton, New Jersey 08541-0001
mrosenfeld@ets.org

Coeditor
F. Jay Breyer, Ph.D.
Thomson Prometric

150 activity statements were distributed across two survey forms (85 activities each) with 20 common to both forms. Also, several demographic questions were asked to collect background data on the sample. A total of 12,000 practitioners were sampled. The WEB group and the MAIL group was each composed of 6,000 practitioners. Within

these calibrations are used to weight the activities. It follows that if there are no differences in the activity calibrations, the weights will be the same across data collection methods and therefore the test plan specifications will be the same. Therefore, the calibrations derived from these two data collection strategies were compared. As an additional step, the manner in which the respondents used the rating scale categories was examined.

. The usage of the frequency rating scale across modes of data collection was nearly identical (Figure 1). Both samples produced virtually identical category probability curves with the thresholds between categories occurring at the same locations. This indicates that both samples understood the meaning of the rating scale categories in similar ways and used them in similar ways. In addition, Figure 2 demonstrates that the relative frequency of the activities were identical across the two modes of data collection.

Figure 1. Category Probability Curves for the Frequency Rating Scale for the MAIL and WEB Samples.

Figure 2. Frequency Calibrations: MAIL vs. WEB.

. The usage of the priority rating scale across modes of data collection was nearly identical (Figure 3). Both samples produced virtually identical category probability curves with the thresholds between categories occurring at the same locations. This indicates that both samples understood the meaning of the rating scale categories in similar ways and used them in similar ways. In addition, Figure 4 demonstrates that the relative priority of the activities were identical across the two modes of data collection.

Figure 3. Category Probability Curves for the Priority Rating Scale for the MAIL and WEB Samples

Figure 4. Priority Calibrations: MAIL vs. WEB.

Because the calibrations for the activities were stable across modes of data collection for both the frequency and priority scales, it follows that both sets of calibrations must yield comparable test plans when the same procedures are used to weight the activities.

Discussion

For this set of data, there were differences in the response rates. MAIL experienced a significantly higher return rate. Although the response rates were adjusted for bad email addresses, it is not known how many email invitations were blocked due to spam filters or sent to email addresses that still existed, but were not used. This may have contributed to this difference. Based on the three reported demographic questions, the two samples appeared to be similar. In one case (months since graduation) they were statistically different, but the practical difference of four months did not seem substantial.

The use of the Rasch model produced very comparable activity calibrations for both the frequency and the priority rating scales. This is not surprising. The real advantage of the Rasch model is that the relative difficulty of the activities is independent of the amount of the latent trait in the sample. Of course if there is one latent trait underlying the responses of one group and a different latent trait underlying the responses of the second group, then it is unlikely that activities will have similar calibrations.

Please note that when using a method like the Spray and Huang (2000) procedure, the importance of the role delineation study that identifies the activities and classifies them into categories should not be taken lightly. The number of activities in each category strongly drives the weighting of the test plan. If the activities in one content area are numerous because they are broken out in very specific detail and in another content area the activities are few because the are stated in a more vague or global way, then the resulting test plan will be over weighted in the first area and under weighted in the other area. It is better to have this issue resolved in the role delineation study, rather than trying to correct it afterward.

Although the results support using WEB to collect practice analysis data for the NCLEX-RN, the results may not generalize to other professions or even to the population of licensed practical nurses (LPN) or vocational nurses (VN). An additional study is underway to examine whether the results also generalize to the LPN/VN population.

Finally, the cost needs to be considered. The estimate for printing and mailing the pre-notice, survey and follow-up reminders was approximately $6.50 per respondent. Given the 6,000 sampled, the cost is approximately $39,000. This figure does not include the management of the returned surveys and the time to scan and QC approximately 1,700 eight page surveys. Including those aspects as well, it is possible that the cost approaches $50,000 for MAIL. The time spent creating an email list, the electronic survey and the QC of the data is much less for the WEB. In addition, printing and mailing costs do not exist. The estimate for the setup and management cost of the WEB is less than $10,000.

Given the cost difference and the fact that the resulting test specifications were comparable, it may be argued that cost is too high to justify a MAIL survey.

References

Andrich, D. (1978). A rating formulation for ordered response categories. , 43, 561-573.

Spray, J. A. & Huang, C. (2000). Obtaining test blueprint weights from job analysis surveys , 37 (3), 187-201.

Wendt, A. & O'Neill, T. (2006). 2005 RN Practice Analysis: Linking the NCLEX-RN® Examination to Practice. Chicago: National Council of State Boards of