

# CLEAR Exam Review

Volume XVII, Number 2

Summer 2006

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](mailto: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.

---

**Editorial Board:** Janet Ciuccio, American Psychological Association; Steven Nettles, Applied Measurement Professionals; Jim Zukowski, Texas State Board of Dental Examiners.

#### Coeditor

Michael Rosenfeld, Ph.D.  
Educational Testing Service  
Princeton, New Jersey 08541-0001  
[mrosenfeld@ets.org](mailto:mrosenfeld@ets.org)

#### Coeditor

F. Jay Breyer, Ph.D.  
Thomson Prometric  
2000 Lenox Drive, Lawrenceville, NJ 08648  
[jay.breyer@thomson.com](mailto:jay.breyer@thomson.com)

## Contents

### From the Editors 1

*F. Jay Breyer, Ph.D.*

*Michael Rosenfeld, Ph.D.*

### Columns

#### Abstracts and Updates 2

*George T. Gray, Ed.D.*

#### Technology and Testing 7

*Robert C. Shaw, Jr., Ph.D.*

### Articles

#### Comparability of Practice Analysis 9

Survey Results Across Modes of Administration

*Thomas R. O'Neill, Reed Castle, Casey Marks*

#### Variable-Length Computerized 13

Classification Testing with Item Response Theory

*Nathan A. Thompson*

#### Correction 19

## Comparability of Practice Analysis Survey Results Across Modes of Administration

*Thomas O'Neill*

*NCSBN*

*Reed Castle*

*SMT*

*Casey Marks*

*NCSBN*

---

*“This study compares two data collection methods for practice analyses, paper & pencil surveys sent by mail (MAIL) versus web-based surveys (WEB) with email notification.”*

---

When devising a licensing examination, it is crucial that the test publisher be able to demonstrate that the test content is connected to practice in an understandable and logical way. This is traditionally accomplished by conducting a role delineation study and a practice analysis survey. The role delineation study identifies the activities that should be included in the practice of the profession, while the practice analysis survey assesses the frequency with which these activities are being performed and how critical it is to correctly perform them. This study compares two data collection methods for practice analyses, pencil & paper surveys sent by mail (MAIL) versus web-based surveys (WEB) with email notification. If the results are comparable, then the more efficient and less expensive strategies could be employed without risk to biasing the results.

As a service to its members, the National Council of State Boards of Nursing (NCSBN) develops, maintains, and administers two nurse licensing examinations, the NCLEX-RN® and NCLEX-PN®. These examinations are based upon large-scale practice analyses that are conducted once every three years. The following study was conducted on data collected from the 2005 RN Practice Analysis (Wendt & O'Neill, 2006). Recently NCSBN has explored the feasibility of using web-based surveys to replace paper surveys sent by mail in the hopes that web-based surveys would produce comparable or better results in less time and with less money.

### Practical Considerations

When mailing surveys the sequencing of the events is important. The availability of slots

of time from printers and mailing houses is subject to fluctuations. The merging of the respondents' addresses with envelopes, surveys, postcards, etc. must be carefully thought out. The cost of postage increases with the number of pieces mailed, and typically the data entry costs for paper surveys is higher than for web-based surveys. While the delivery and responses are not instantaneous for MAIL surveys, WEB surveys can be easily analyzed and reviewed in real time. Also, they can be completed in a shorter amount of time and, if needed, can be modified without re-printing other surveys. Intuitively, it would seem that the advantages of WEB over MAIL are so overwhelming that it hardly warrants a study.

Yet, there are some potential limitations associated with WEB. Steps must be taken to prevent people who were not solicited from responding and to ensure that respondents who are solicited get only one “vote.” Assigning unique user IDs and passwords to each respondent usually resolves these problems. Without these restrictions, unsolicited responses may corrupt the sample. Furthermore, it is crucial that current email addresses are available for the population of interest and that there are no important differences in the population between those that use email and those that do not.

### **Method**

#### Study Design

In a role delineation study, 150 activities were identified as comprising the scope of practice. Due to the large number of activities statements on the survey, two forms of the survey (Forms A & B) were used. These

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 each group, half was randomly assigned to Form A and the other half to Form B. Each activity was assessed using two rating scales, frequency (0-5) and priority (1-4). The MAIL sample was a random sample from all newly licensed NCLEX candidates stratified by state or jurisdiction. The WEB sample was selected from the remaining NCLEX candidates who had also submitted their email address. All participants were newly licensed nurses.

#### Mailing Procedure

A pre-survey notice postcard was sent to all 6,000 MAIL participants and emailed to the 6,000 WEB participants. One week later, the survey was sent to the MAIL cohort and a web link to the online survey was emailed to the WEB cohort. Participants had approximately six weeks to complete the survey. During this time, three reminder postcards were sent to the MAIL cohort and three reminder emails were sent to the WEB cohort. As an incentive, all respondents were entered into a drawing and ten randomly drawn winners were given a cash award. Finally, all respondents were sent a letter of recognition for their contribution and time.

#### Research Questions

This study attempts to address three research questions. The first question was, are the response rates different by survey administration mode? The second question was, are there demographic differences across respondents by mode of administration? Finally, would the data from the different modes of survey administration produce different recommended test specifications?

### **Results**

#### Response Rate

One of the potential disadvantages to using a WEB method is that there is the risk of incorrect or outdated email addresses. For the MAILs, there were 263 surveys that were returned for incorrect addresses or due to the fact that the nurse was no longer practicing. For the WEBs, there were a total of 1,116 survey invitations returned due to incorrect email addresses or due to the fact that the nurse was no longer practicing. The corrected response rates were 30% for the MAILs and 23% for the WEBs when only those people who actually received surveys were considered. A z-test of proportions was calculated to determine if there was a difference. A statistically significant difference was observed at the 1% level ( $z = 8.10$ ).

#### Demographic Differences

Differences in demographics were evaluated to determine if the responding samples reflected the same population of practitioners. While there were many demographic variables collected, only three variables are reported in this paper. Tables 1 and 2 present the observed response rates for population density and the type of facility, as well as the average number of months since graduation.

**Table 1. Practice Setting Area**

Practice Setting (Population Density)	MAIL response rate	WEB response rate
Urban	62.2%	63.2%
Suburban	25.5%	27.5%
Rural	12.2%	9.5%

#### Practice Setting (Type of Facility)

Practice Setting (Type of Facility)	MAIL response rate	WEB response rate
Hospital	85.2%	90.0%
Long-term care	7.8%	3.8%
Community/ Ambulatory	5.4%	4.6%
Other	1.5%	1.6%

**Table 2. Months Since Graduation**

Survey mode	Mean	SE
WEB	9.91	1.24
MAIL	14.41	0.42

Practice setting saw minor differences across the two modes of administration. Proportionally, rural respondents were better represented, by almost 3%, with the MAIL mode. There was a minor difference, albeit statistical, in the months since graduation between survey administration modes. Those responding to the MAIL reported a statistically significant higher number of months since graduation. The difference was approximately 4.5 more months for MAIL respondents. Facility type saw similar response rates in the community/ambulatory and other categories and minor differences in hospital and long-term care settings. Facility type had two categories that were similar and two categories (hospital and long-term care) that had a four to five percent difference.

#### Expected Impact on Test Specifications

Although there will always be minor differences among different subsets of the sample, the important question is, will the resulting recommendations for the test plan specifications be different under these two data collection methodologies? NCSBN uses a variation of Spray and Huang's (2000) Rasch-based procedures for deriving test plan specifications. In these procedures, the activities are calibrated using Andrich's (1978) rating scale model and

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.

*Frequency.* 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.

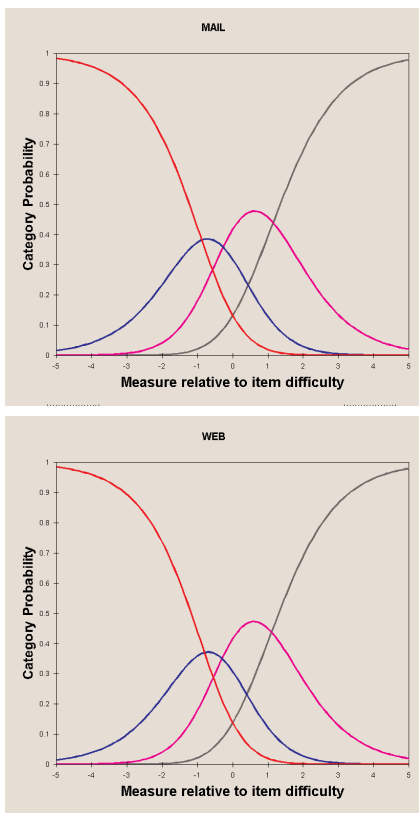
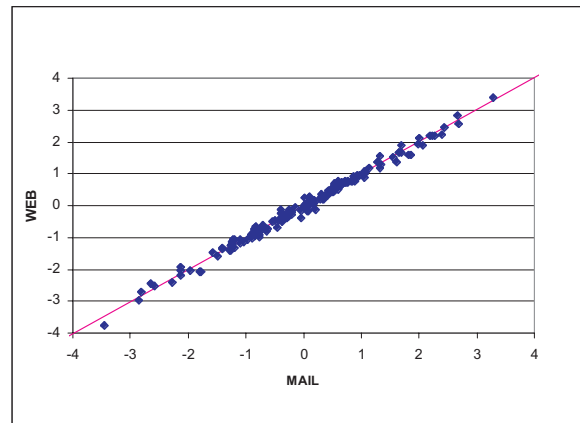


Figure 2. Frequency Calibrations: MAIL vs. WEB.



*Priority.* 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

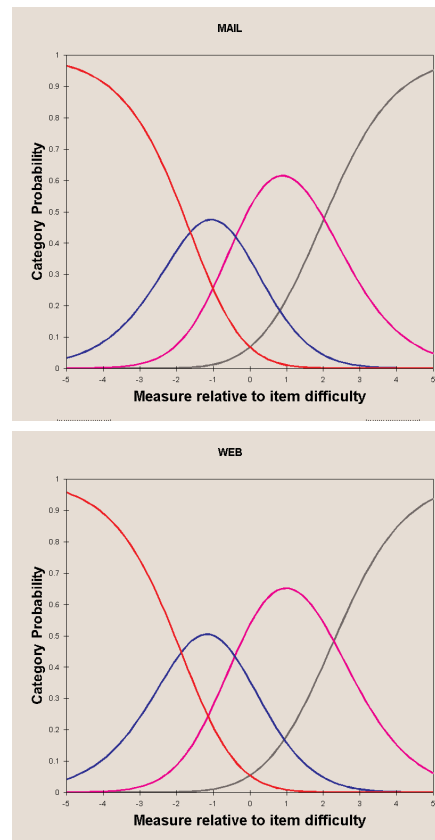
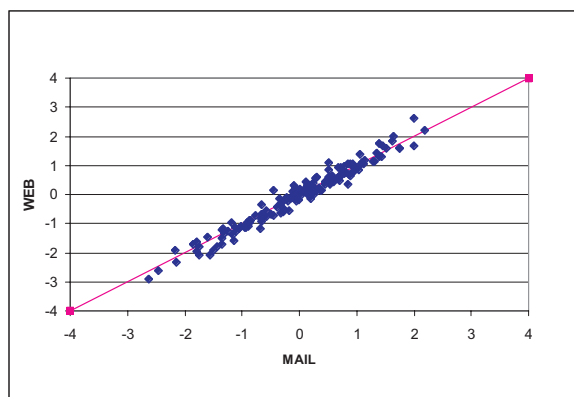


Figure 4. Priority Calibrations: MAIL vs. WEB.



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 they 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. *Psychometrika*, 43, 561-573.
- Spray, J. A. & Huang, C. (2000). Obtaining test blueprint weights from job analysis surveys. *Journal of Educational Measurement*, 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 Nursing.