CLEAR Exam Review

Volume XIX, Number

A Journal

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 200, Lexington, KY 40502.

Editing and composition of this journal have been written by 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 Classified (e.g., position vacancies) for CER may be reserved by contacting Janet Horne 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
Applied Measurement Professionals

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

Coeditor
F. Jay Breyer, Ph.D.
Prometric
2000 Lenox Drive
Lawrenceville, NJ 08648
jay.breyer@prometric.com

CLEAR Exam Review

Contents

FROM THE EDITORS 1

F. Jay Breyer, Ph.D.
Michael Rosenfeld, Ph.D.

Abstract

Advances in computer-based testing and Item Response Theory have created opportunities for the National Nursing Licensure Examination For Registered Nurses (NCLEX-RN®) to explore innovative items. This article compares traditional multiple-choice items with some innovative formats such as fill-in-the-blank calculation items, fill-in-the-blank ordered response items and multiple response items. Using two experimental datasets that were created from two time periods when the innovative items were pretested, items were calibrated using the Rasch (1PL) measurement model. Results of this study indicate that innovative items offer measurement properties that are comparable to or at times better than traditional multiple-choice items.

Introduction

Over a decade ago (1994) the U.S. National Nursing Licensure Examinations (NCLEX-RN®) moved from paper-and-pencil format using standard, four-option multiple-choice questions (MCQs) to Computerized Adaptive Technology (CAT) using that same item format. At that time, it was postulated that computers have the potential to assess new skills and abilities that have been difficult or extremely expensive to measure via traditional testing formats (McHenry & Schmitt, 1994). Innovations in computer-based testing include item types with features that include sound, graphics, animation and video integrated into the item stem, response options or both. In addition, use of Item Response Theory (IRT) has allowed the creation of measurement scales that are independent of the particular sample of people or test items used to create the scales (Lord & Novick, 1968; Lord, 1980). Furthermore, the use of IRT has facilitated the introduction of CAT for testing programs. With the introduction of CAT and innovative items, one research issue that is important to address is whether the innovative item types behave in ways that are compa ncurest r-kBiitemthy

When introducing new item formats, one major concern is with dimensionality. A second major concern is with model-data fit. These are somewhat related issues, but it is possible for items to measure one major factor (dimension) and yet fit the models slightly differently. Large systematic problems with model-data fit ("misfit") might indicate some problem with dimensionality. This paper will focus on issues of model-data fit rather than issues of dimensionality because the design for this study, using current pretest data, did not produce data that would permit a reliable look at dimensionality. Thus, the purpose of this paper is to examine whether the innovative items are similar to MCQs in terms of their item statistical characteristics.

Methodology

To approach the issue of model-data fit, large datasets that are available from ongoing pretesting of the alterR…art

Multiple Response (MR)

While traditional MCQs allow the examinee to select a response from a list of four options, the multiple response innovative item is a variant on this item type that allows the examinee to choose one or more of the options provided (e.g., options 1,3, and 6). Figure 3 is an example of a Multiple Response item. This item format is used without cueing the examinee to the actual number of correct responses. Additionally, this format requires that the examinee have the ability to discriminate from a list of important content which has implications for examinees ability to think critically (Jodoin, 2003). Within nursing content, this item type is intended to identify the examinee's ability to consider all possibilities in providing client care in a given situation. Depending on the phrasing of the content in the item, an examinee may be required to discriminate between non-mutually exclusiwn,

p

datasets, the MCQ items flagged for “misfit” based on the Infit index ranged from 15.1% to 28.4%, and the MC items flagged for misfit based on the Outfit index rangedg, f

significant increase in the number of candidates running out of time on the exam (NCSBN, 2006). Additionally, innovative items are being introduced into the examination in a measured way. As new item formats are introduced, further investigation will be needed to ensure that the items offer measurement properties that are comparable to if not better than multiple-choice items.

References

Ackerman, P. (2003). Cognitive ability and non-ability trait determinants of expertise. *Educational Researcher*, 32(8), 15-20.

Bejar, I. I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. *Journal of Educational Measurement*, 17, 283-296.

Bond, Trevor & Fox, Christine (2001). Applying the rasch model. *Fundamental measurement in the human sciences.* Chapter 12, The question of model fit. Mahwah, NJ: Erlbaum.

Douglas, Graham (1982). Issues in the Fit of Data to Psychometric Models. *Education Research and Perspectives* 9:1, 32-43.

Haladyna, Thomas M. (1997). *Writing test items to evaluate higher order thinking.* Allyn & Bacon. Needham Heights, Mth/[PtQ]Tu/[PNQ]T/[PJQ]TZb]T/[PNQ]T/[PJQbHVRUZb]T/[PHQ]T/[Hbcu/[PhQ]T/[PtQ]Yq]T/[PaQ]T/[PtQ]T/[PeQ]T/erA5A5a thevatssetem