ED 476 422   TM 034 924

AUTHOR   Hertz, Norman R.; Chinn, Roberta N.
TITLE   Effects of Item Exposure for Conventional Examinations in a
                Continuous Testing Environment.
PUB DATE   2003-04-00
NOTE   16p.; Paper presented at the Annual Meeting of the National
                Council on Measurement in Education (Chicago, IL, April 22-
                24, 2003).
PUB TYPE   Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE   EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS   *Computer Assisted Testing; Item Banks; Item Response Theory;
                *Licensing Examinations (Professions); Social Workers; *Test
                Items; Therapists
IDENTIFIERS   *Item Exposure (Tests)

ABSTRACT
                This study explored the effect of item exposure on two
conventional examinations administered as computer-based tests. A principal
hypothesis was that item exposure would have little or no effect on average
difficulty of the items over the course of an administrative cycle. This
hypothesis was tested by exploring conventional item statistics and Rasch
estimates of ability and difficulty of four separate groups of candidates who
took a licensing examination in a continuous testing environment over the
course of a 6-month administration cycle. Subjects were 1,001 candidates for
a state license in clinical social work (LCSW) and 1,660 candidates for
licensure as a marriage and family therapist (MFT). Taken together, results
suggest that even if candidates took the examinations later in the cycle,
there was no clear indication that information obtained form candidates who
took the test early in the cycle improved performance. The most important
implication for small testing programs is that they can enjoy the benefits of
computer administration without having large item pools and candidate
populations or using computer adaptive delivery systems. (Contains 4 figures,
2 tables, and 22 references.) (SLD)

ED 476 422

TM034924

Effects of Item Exposure for

Conventional Examinations in a

Continuous Testing Environment

Norman R. Hertz and Roberta N. Chinn

HZ Assessments[1]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

● Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

N. R. Hertz

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

Presented at the annual meeting of the National Council on Measurement in Education, Chicago,

Illinois, April 2003.

[1] Author note: Correspondence should be addressed to Norman R. Hertz, Ph. D., HZ Assessments, 177 Arbuckle
Avenue, Folsom, California 95630. Electronic mail may be sent to hzmail@attbi.com.

BEST COPY AVAILABLE

A. OBJECTIVES OF THE INQUIRY

Many small licensing and certification programs are attracted to computer-based testing, rather than computerized adaptive testing, as a medium for test delivery. Computer-based testing usually involves fixed forms offered on a computer for a specified period of time. There are many advantages of computer technology: better administrative control, randomized ordering of items for each candidate, flexible candidate scheduling, and immediate delivery of scores to candidates (Way, 1998). Despite the efficiencies of computer technology, there is concern about item security, particularly if a fixed form is available for several months. Thus, the concern underlying item exposure is that candidates who have prior knowledge of examination content will achieve higher scores than candidates with no prior knowledge.

Most of the literature regarding item exposure is concentrated on computer adaptive testing. Several studies addressed the issue of item exposure in a computerized adaptive testing environment in terms of the effect of item exposure on examinee scores and different item selection strategies such that as item exposure increases, measurement precision decreases (Hale, Angelis, & Thibodeau, 1983; O'Neill, Lunz, & Thiede, 2000; Pastor, Dodd, & Chang, 2002; Stocking & Lewis, 1998; Stocking & Swanson, 1998; Way, 1998). These authors propose a number of techniques to control item exposure yet maintain measurement precision. There are two main approaches to control item exposure in computerized adaptive testing. randomized item selection (e. g., Bergstrom, Lunz, & Gershon, 1992; McBride & Martin, 1983) and conditional item selection (e, g, Stocking and Lewis, 1998; Sympson & Hetter, 1985). Alternative procedures such as the stratified design procedure attempt to control item exposure by stratifying the item selection mechanism (Chang & Ying, 1999).

Other studies have examined the effect of administering the same performance-based examination to different cohorts of students over several days or weeks resulted in no consistent change across student cohorts (Colliver, Barrows, Vu, Verhulst, Mast & Travis, 1991; Stillman, Haley, Sutnick, Philbin, Smith, O'Donnell, & Pohl, 1991; Rutala, Witzke, Leko, Fulginiti, & Taylor, 1991). Some studies indicated that access to test information from students early in the cycle did not affect the test scores of students who took the examination later in the cycle (Skakun, Cook, & Morrison, 1992; Swartz, Colliver, Cohen, & Barrows, 1995).

In sum, the findings from the aforementioned studies are mixed. Two studies note that there can be differences in testing outcomes that are statistically but not practically significant (O'Neill et al., 2000; Stocking, Ward, & Potenza, 1998). Nonetheless, there appear to be a variety of factors that affect test outcomes when items are exposed throughout a computerized testing cycle. There are psychometric factors: item structure and test specifications (Stocking & Lewis, 1998), examinee ability (O'Neill et al., 2000), test length and item pool size (Pastor, et al., 2002; Reveulta & Ponsoda, 1998; Way, 1998), type of examination such as standardized patient examinations (Macmillan, De Champlain, & Klass, 1999), and exposure rate and percent of item overlap (Stocking, et al., 1998; Way, 1998). Non-psychometric factors could also affect test outcomes such as the specific recall effect, or an increase in performance due to recall of specific questions and answers (Hale et al., 1983).

The effects of either psychometric or non-psychometric factors have yet to be studied in terms of fixed forms administered in a continuous testing environment. Item exposure in a continuous testing environment is of particular significance because items are available over a relatively long period of time during an administrative cycle, and 30% of the items may be reused as anchor items from one form to the next for a subsequent cycle. The important question to be

asked is if the same set of items were available throughout a cycle, would candidate performance increase from the beginning of the cycle to the end of the cycle? A related question to be asked is what could account for changes in the percentage passing over the course of a cycle? The answers to these questions is of obvious interest to small testing programs, whose item banks may be small and whose candidate pools are not sufficient to use computer-adaptive testing delivery systems.

In the present study, the effect of item exposure on two conventional examinations administered as computer-based tests was explored. A principal hypothesis was that item exposure would have little or no effect on average difficulty of the items over the course of an administrative cycle. There are several factors that underlie this hypothesis. First, only small changes in the percent passing had been observed in previous administrations. Second, the examinations were relatively long (175 items each). Third, the content of the items spanned a broad range of topics in their respective content specifications, e.g. assessment, therapeutic interventions, legal and ethical responsibilities, etc. Fourth, each candidate received a different random order of items. Finally, many of the items required the candidates to apply their clinical knowledge and training to a case scenario, a format in which it is difficult to memorize specific bits of information.

The hypothesis was tested by exploring conventional item statistics and Rasch estimates of ability and difficulty of four separate groups of candidates who took a licensing examination in a continuous testing environment over course of a 6-month administration cycle.

B. SOURCES OF INFORMATION PRESENTED

Subjects. Subjects were 1,001 candidates for a state license in clinical social worker (LCSW) and 1,660 candidates for a state license as a marriage and family therapist (MFT).

Examinations. The examinations were administered as licensing examinations for LCSWs and MFTs. All examinations were multiple-choice and contained 175 items. The examinations

used in the present study were: two forms of an LCSW examination and two forms of an MFT examination. The amount of overlap of items between forms was 30%. Cut scores for each examination were established with a modified Angoff procedure.

Computer-based testing environment. The examinations were administered as fixed, linear forms in a continuous testing environment. There were four 6-month periods of administration: November 16, 2000-May 15, 2001; May 16, 2001-November 15, 2001; January 2, 2001-June 29, 2001; and, June 30, 2001-December 31-2001. Candidates were only allowed to take the examination once within each 6-month period a candidate. Candidates who failed the examination during a given 6-month period could retake the examination when another form became available in the following 6-month period.

C. METHODS

Design. For each of the four examinations, data were partitioned in three 2-month periods to create a means to analyze changes within an administrative cycle. For example, the November 1, 2000 – May 15, 2001 administration of the LCSW examination, data were divided into three 2-month periods (first, second, third).

Data analyses. Conventional item analyses were performed on each examination to calculate conventional item statistics for each 2-month period and for the entire 6-month period. There were not sufficient numbers of candidates who sat for the examination to meet the criteria for two-parameter or three-parameter models (e. g., Hulin, Lissak, & Drasgow, 1983); however, estimates could be obtained with a Rasch model (e. g., Green, Bock, Humphreys, Linn, & Reckase, 1984; Wright & Stone, 1979). In the present study, separate Rasch analyses were performed to obtain mean difficulty estimates and mean ability estimates for each 2-month period and for the entire 6-month period.

D. RESULTS AND CONCLUSIONS

Tables 1 and 2 summarize the mean test scores, ability estimates, and difficulty estimates for both examinations. Overall, the effect of item exposure was slight. For all four examinations, there was no systematic increase over the 6-month period in terms of the percent passing or coefficient alpha. For the conventional item analyses, the mean test scores and standard deviations of scores changed only slightly (range of mean scores: 1.22 to 3.60 points; range of standard deviations: 1.13 to 2.38 points) during each 2-month period in an administrative cycle.

For the Rasch analyses where difficulty was centered on zero, there were only slight changes in the mean and standard deviations of ability estimates each 2-month period for a given examination (see Figures 1, 2, 3, 4). When ability was centered on zero, there were only slight changes in mean and standard deviations of the difficulty estimates.

Taken together, the results suggest that even if candidates took the examination later in the cycle, there was no clear indication that information obtained from candidates who took the examination early in the cycle improved performance.

E. EDUCATIONAL IMPORTANCE OF THE STUDY

While the results of the present study are preliminary, the study provides information on how item exposure functions in a realistic context of licensure examinations administered in a continuous testing environment. If a fixed form is administered in a continuous testing environment, the performance of candidates did not increase consistently over time. The most important implication for small testing programs is that they can enjoy the benefits of computer administration without having large item pools and candidate populations or using computer adaptive delivery systems.

REFERENCES

Bergstrom, B. A., Lunz, M. E., & Gershon, R. C. (1992). Altering the level of difficulty in computer adaptive testing. Applied Measurement in Education, 5, 137-149.

Chang, H. & Ying, Z. (1999). A stratified multistage computerized adaptive testing. Applied Measurement in Education, 23, 211-222.

Colliver, J. A., Barrows, H. S., Bu, N. B., Verhulst, S. J., Mast, T. A., & Travis, T. A. (1991). Test security in examinations that use standardized-patient cases at one medical school. Academic Medicine, 5, 279-282.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 21, 347-360.

Hale, G. A., Angelis, P. J., & Thibodeau, L. A. (1983). Effects of item disclosure on performance on the Test of English as a Foreign Language. Language Learning, 33, 449-464.

Hulin, C. L., Lissak, F. & Drasgow, C. K. (1983). Item response theory: Applications to psychological measurement. Homewood, IL: Down Jones-Irwin.

Macmillan, M., K., De Champlain, A. F., & Klass, D. J. (1999). National security threats to licensing examinations using standardized patients. Academic Medicine, 74 (supplement), S55-S57.

McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In: D. J. Weiss (Ed.), New horizons in testing (pp. 223-236). New York: Academic Press.

O'Neill, T., Lunz, M. E., & Thiede, K. (2000). The impact of receiving the same items on consecutive computer adaptive test administration. Journal of Applied Measurement, 1, 131-151.

Pastor, D. A., Dodd, B. G., & Chang, H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. Applied Psychological Measurement, 26, 147-163.

Reveulta, J., & Ponsoda, V. (1998). A comparison if item exposure control methods in computerized adaptive testing. Journal of Educational Measurement, 35, 311-327.

Rutala, P. J., Wizke, D. B., Leko, E. O., Fulginiti, J. V., & Taylor, P. H. (1991). Archives of Internal Medicine, 151, 541-544.

Skakun, E. N., Cook D. A., & Morrison, J. C. (1992). Test security on sequential OSCE and multiple choice examinations. In: R. M. Harden, I. R. Hart, & H. Mulholland, (Eds.). Approaches to the assessment of clinical competence (pp. 711-718). Norwich, U. K.: Page Brothers.

Stillman, P. L., Haley, H. A., Philbin, M. M., Smith, S. R., O'Donnell, J., & Pohl, H. (1991). Is test security an issue in multistation clinical assessment?—A preliminary study. Academic Medicine, 66 (supplement), S25-S27.

Stocking, M. L., & Lewis, C. (1993). A method for severely constrained item selection in adaptive testing. Applied Psychological Measurement, 17, 277-292.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. Journal of Educational and Behavioral Statistics, 23, 57-75.

Stocking, M. L., & Swanson, L. (1998). Optimal design of item banks for computerized adaptive tests. Applied Psychological Measurement, 22, 271-279.

Stocking, L., Ward, W. C., & Potenza, M. T. (1998). Stimulating the use of disclosed items in computerized adaptive testing. Journal of Educational Measurement, 35, 48-68.

Swartz, M. H., Colliver, J. A., Cohen, D. S., & Barrows, H. S. (1995). The effect of deliberate, excessive violations of test security on a standardized-patient examination: An extended analysis. In: A. I. Rothman, & R. Cohen (Eds.), Proceedings of the Sixth Ottawa Conference on Medical Education, 1994, July 26-29, Toronto (pp. 280-284). Toronto, Ontario, Canada: University of Toronto Press

Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. Proceedings of the 27th Annual Meeting of the Military Testing Association (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.

Way, W. (1998). Protecting the integrity of computerized testing item pools. Educational Measurement: Issues and Practice, 17, 17-27.

Wright, B. D. & Stone, M. H. (1979). Best test design. Chicago, MESA.

Table 1

Conventional and Rasch Analyses for LCSW Examinations (175 items each)

November 16, 2000 – May 15, 2001 Administration [a]

	n	Conventional Statistics				Rasch Centered on Ability		Rasch Centered on Difficulty	
		Percent Passing	M_{test}	SD_{test}	Alpha	$M_{difficulty}$	$SD_{difficulty}$	$M_{ability}$	$SD_{ability}$
First 2-month period	126	29.37%	111.07	14.69	.85	-1.49	2.27	.68	.44
Second 2-month period	151	29.80%	110.25	14.19	.83	-1.48	2.22	.65	.42
Third 2-month period	189	38.10%	112.79	12.03	.77	-1.86	2.61	.73	.36
Total 6-month period	466	33.04%	111.50	13.54	.82	-1.62	2.32	.69	.40

[a] Cut score is 119

May 16, 2001 – November 15, 2001 Administration [b]

	n	Conventional Statistics				Rasch Centered on Ability		Rasch Centered on Difficulty	
		Percent Passing	M_{test}	SD_{test}	Alpha	$M_{difficulty}$	$SD_{difficulty}$	$M_{ability}$	$SD_{ability}$
First 2-month period	214	64.95%	121.58	12.48	.81	-2.25	2.47	1.01	.41
Second 2-month period	137	57.66%	117.98	14.15	.85	-1.88	2.21	.89	.44
Third 2-month period	184	57.07%	118.13	13.22	.83	-2.01	2.46	.91	.42
Total 6-month period	535	60.37%	119.47	13.29	.83	-2.05	2.34	.94	.43

[b] Cut score is 119

Table 2

Conventional and Rasch Analyses for MFT Examinations (175 items each)

January 2, 2001 – June 29, 2001 Administration [a]

		Conventional Statistics				Rasch Centered on Ability		Rasch Centered on Difficulty	
	n	Percent Passing	M_{test}	SD_{test}	Alpha	$M_{difficulty}$	$SD_{difficulty}$	$M_{ability}$	$SD_{ability}$
First 2-month period	281	57.30%	120.74	15.74	.87	-1.93	1.72	.95	.48
Second 2-month period	365	60.00%	121.68	14.61	.85	-2.12	1.85	.98	.45
Third 2-month period	184	54.35%	120.46	14.66	.85	-2.07	1.85	.85	.44
Total 6-month period	830	57.83%	121.09	16.02	.86	-2.04	1.78	.96	.46

[a] Cut score is 120

June 30, 2001 – December 31, 2001 Administration [b]

		Conventional Statistics				Rasch Centered on Ability		Rasch Centered on Difficulty	
	n	Percent Passing	M_{test}	SD_{test}	Alpha	$M_{difficulty}$	$SD_{difficulty}$	$M_{ability}$	$SD_{ability}$
First 2-month period	356	52.53%	117.43	16.15	.87	-1.73	1.83	.94	.48
Second 2-month period	258	58.91%	119.58	13.77	.83	-2.14	2.21	.93	.42
Third 2-month period	216	56.48%	117.20	15.43	.86	-1.83	2.04	.86	.46
Total 6-month period	830	52.65%	118.04	15.29	.86	-1.86	1.96	.88	.46

[b] Cut score is 118

Figure 1. Item by Person Distribution Maps for Three LCSW Examinations - November 16, 2000 – May 15, 2001 Administration
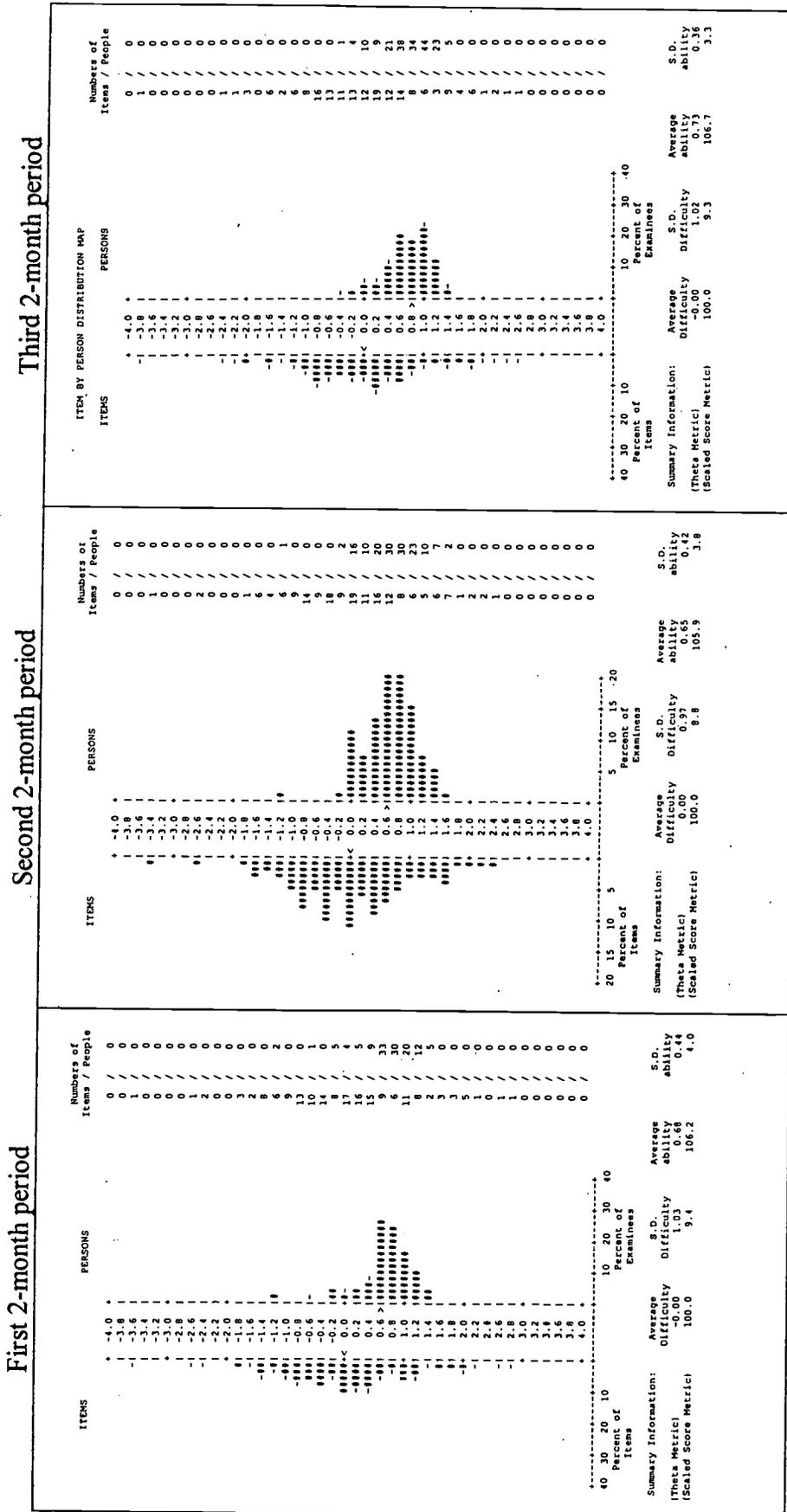
Figure 2. Item by Person Distribution Maps for Three LCSW Examinations - May 16, 2001 November 15, 2001 Administration
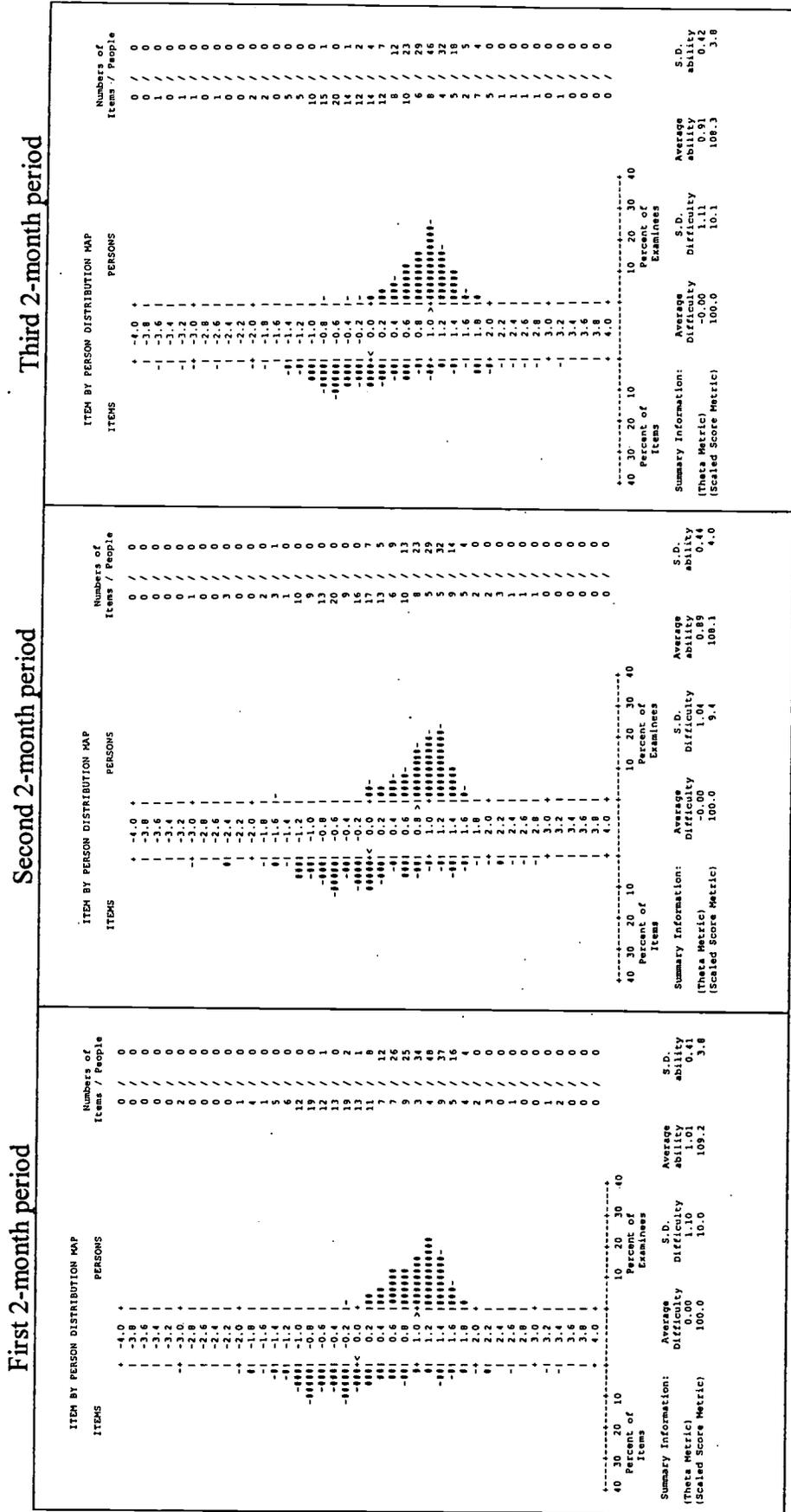
First 2-month period

Second 2-month period

Third 2-month period

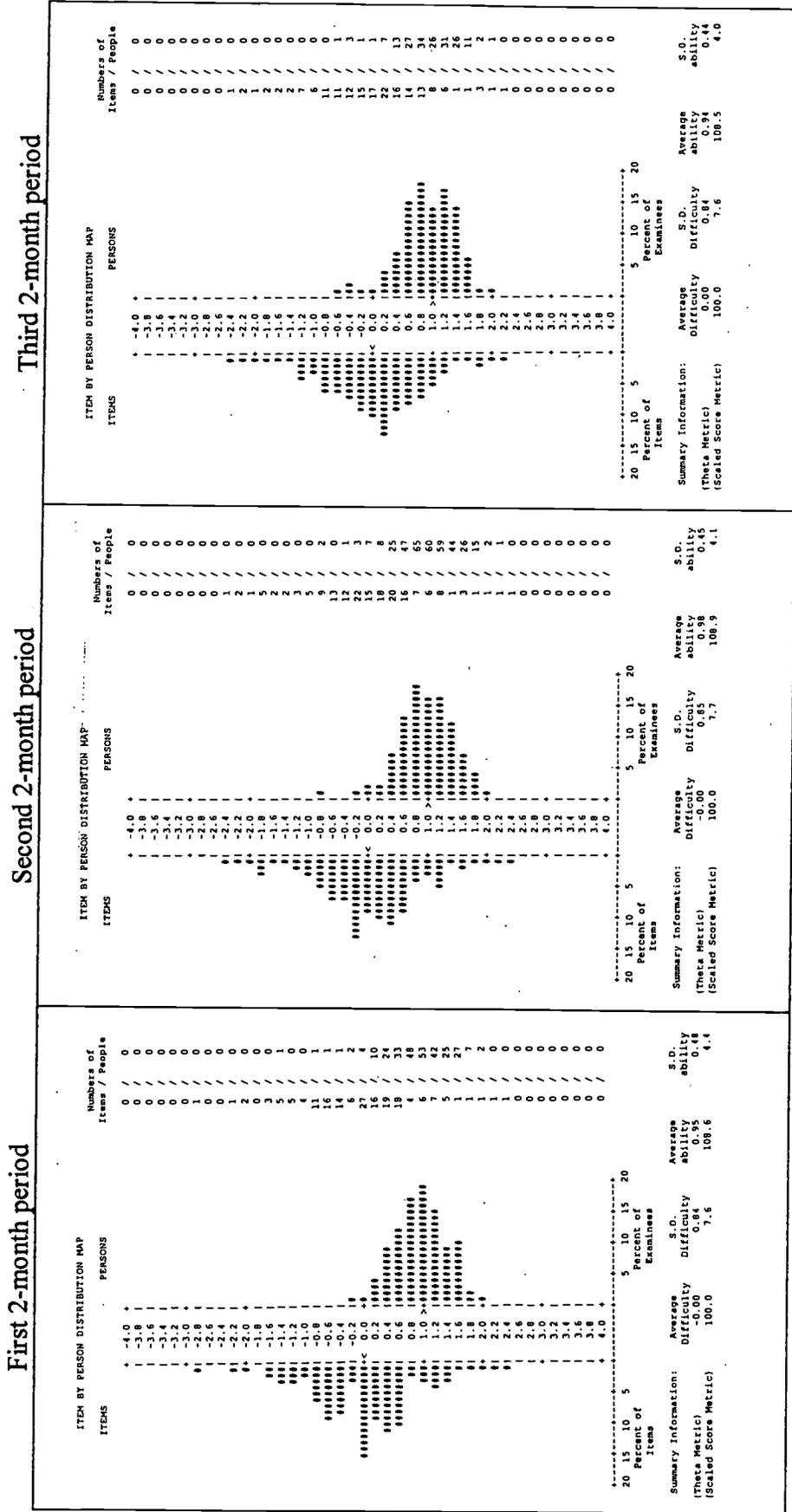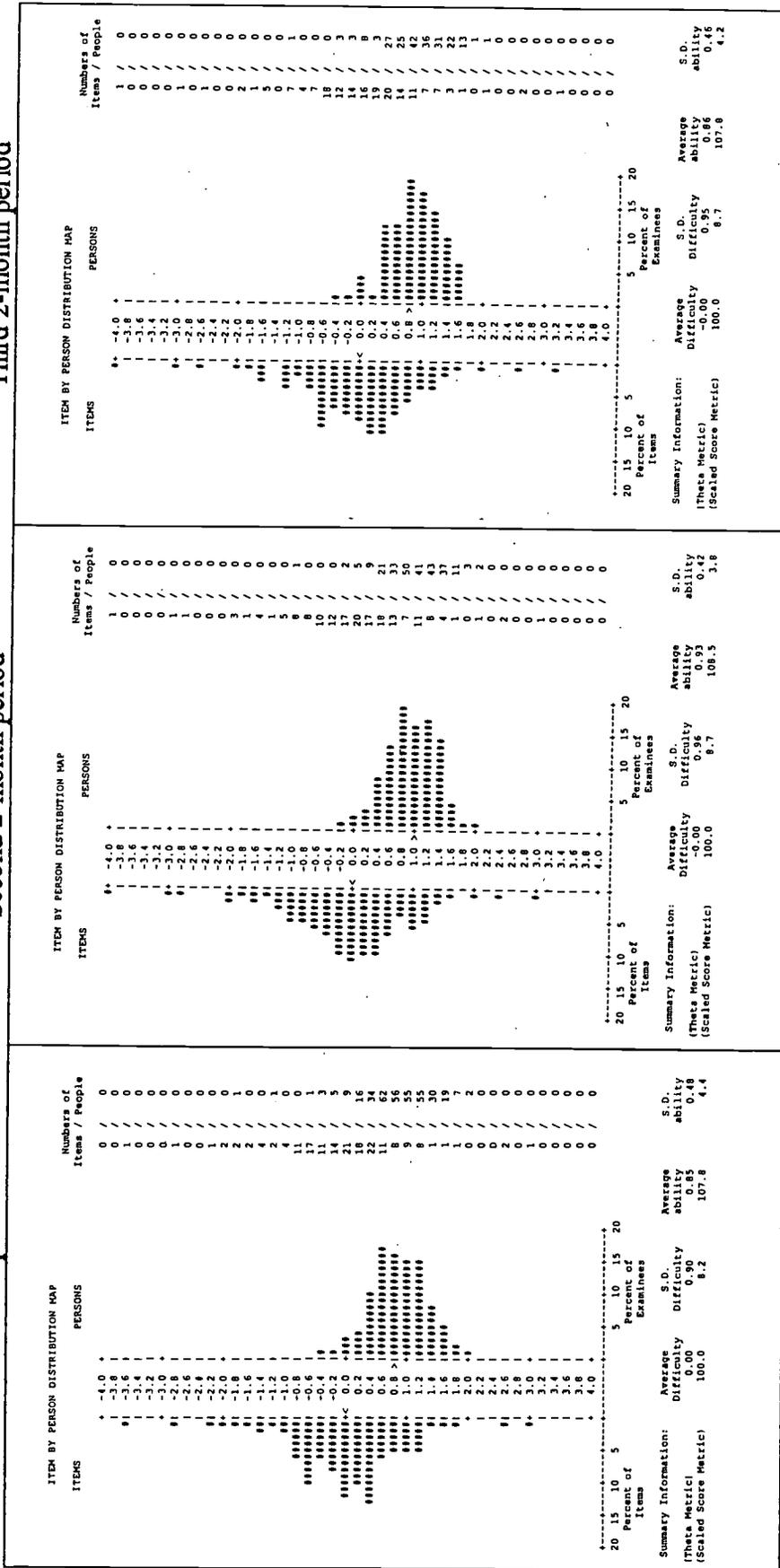Figure 3. Item by Person Distribution Maps for Three MFT Examinations – January 2, 2001 – June 29, 2001 Administration

First 2-month period   Second 2-month period   Third 2-month period

Figure 4. Item by Person Distribution Maps for Three MFT Examinations (June 30, 2001 – December 31, 2001 Administration)

First 2-month period

Second 2-month period

Third 2-month period

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)
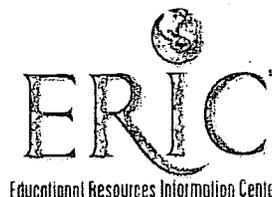
REPRODUCTION RELEASE
(Specific Document)

ERIC
Educational Resources Information Center

I. DOCUMENT IDENTIFICATION:

Title: *Effects of item exposure for conventional examinations in a continuous testing environment*

Author(s): *Norman R. Hertz, Roberta N. Chinn*

Corporate Source: *H2 Assessments*

Publication Date: *April 24, 2003*

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY _____ Sample _____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY _____ Sample _____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY _____ Sample _____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ↑ [X]	Level 2A ↑ [ ]	Level 2B ↑ [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signed: *Norman R. Hertz*

Printed Name/Position/Title: *NORMAN R. HERTZ, PARTNER*

Organization/Address: *H2 ASSESSMENTS, 177 ARBUCKLE AVE FOLSOM, CA 95630*

Telephone: *916.351.9147* FAX: *916.353.1147*

E-Mail Address: *H2MAIL@ATT81.COM* Date: *5-11-2003*

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions