DOCUMENT RESUME

ED 442 874		TM 031 300
AUTHOR	Huyhn, Huynh	
TITLE	On Item Mappings and Statistical Rules for Selecting Binary Items for Criterion-Referenced Interpretation and Bookmark Standard Settings.	
PUB DATE	2000-04-00	
NOTE	10p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 25-27, 2000).	
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Bayesian Statistics; *Criterion Referenced Tests; *Selection; Standards; Test Construction; Test Interpretation; *Test Items	
IDENTIFIERS	*Mapping; *Standard Setting	

ABSTRACT

Item mappings are widely used in educational assessment for applications such as test administration (through test form assembly and computer assisted testing) and for criterion-referenced (CR) interpretation of test scores or scale anchoring. Item mappings are also used to construct ordered item booklets in the CTB/McGraw Hill Bookmark standard setting procedure. Selection rules for mapping the items vary with the purpose of the mapping. The objective of this paper is to categorize various types of item mappings, to describe ways to assess the consequences of a given item selection rule for mapping a binary item, and to provide a general empirical Bayes framework from which specific selection rules can be derived. A comparison is made on the maximum information (MI) rules and those derived from an empirical Bayes (EB) approach. It is noted that the EB rules coincide with the MI rules if the correction for guessing formula is used to extend the EB rules for Rasch and two parameter logistic items to the EB rules for three parameter logistic items. (Contains 13 references.) (Author/SLD)

TM

On Item Mappings and Statistical Rules for Selecting Binary Items for Criterion-Referenced Interpretation and Bookmark Standard Settings

Huynh Huynh
University of South Carolina

Item mappings are widely used in educational assessment for applications such as test administration (through test form assembly and computer assisted testing, CAT) and for criterion-referenced (CR) interpretation of test scores or scale anchoring. Item mappings are also used to construct ordered item booklets in the CTB/McGrawHill Bookmark standard setting procedure. Selection rules for mapping the items vary with the purpose of the mapping. The objective of this paper is to categorize various types of item mappings, to describe ways to assess the consequences of a given item selection rule for mapping a binary item, and to provide a general empirical Bayes framework from which specific selection rules can be derived. A comparison is made on the maximum information (MI) rules and those derived from an empirical Bayes (EB) approach. It is noted that the EB rules coincide with the MI rules if the correction for guessing formula is used to extend the EB rules for Rasch and 2PL items to the EB rules 3PL items.

Introduction

Locating an item on an achievement continuum (*item mapping*) is a well-entrenched process in educational assessment. Applications of item mapping may be found in criterion-referenced (CR) testing (or scale anchoring, Beaton & Allen, 1992; Huynh, 1994, 1998), computer-assisted testing, test form assembly, and in selecting items for ordered test booklets used in the Bookmark standard setting process (Lewis, Mitzel, & Green, 1996). While item response theory (IRT) models such as the Rasch and two-parameter logistic (2PL) models traditionally place a binary item at its location, it has been argued (Huynh, 1998) that such mapping may not be appropriate in selecting items for CR interpretation and scale anchoring.

A paper presented at the annual meeting of National Council on Measurement in Education, New Orleans, April 2000. *The author's address is College of Education, Wardlaw 138, University of South Carolina, Columbia SC 28208. Email: huynh-huynh@sc.edu.*

The purpose of this paper is to describe the three types of item mappings that are often used in applications. Attention will then be focused on the selection of binary items for CR interpretation (or scale anchoring) and for the construction of ordered test booklets in the Bookmark standard setting process. A statistical framework will be provided for assessing the consequences of each selection procedure. Within an empirical Bayes context, specific selection rules will be formulated and compared with the maximum information (MI) rules derived by Huynh (1994, 1998).

Three Types of Item Mapping

There are different ways to map an item on an achievement continuum and the statistical rule for mapping depends on the specific purpose of the item mapping. A extensive discussion of the selection rules may be found in Huynh (1998). Although item mappings can be categorized in many different ways, there are three major types that are being used in many applications. These three types differ in terms of context, application, statistical and psychometric framework, and implications. They are described as follows.

Type 1: Item Mapping for Ability Estimation

In the context of ability estimation, items are typically chosen to match the ability of the examinee and an ability estimate is obtained from the selected items. It is here that the Fisher information plays a major role. Items that minimize the (asymptotic) standard error of the estimated ability are those that maximize the Fisher information. Details about item mapping for this situation may be found in traditional textbooks on IRT such the one by Hambleton and Swaminathan (1985). Item selection in this case aims only at an accurate estimation of the examinee's ability and skip over all questions about what the examinee can be expected to accomplish at this level of performance. In general, let a, b and c be the item parameters of a three-parameter logistic (3PL) item. The probability of answering the item correctly is given as

$$P(\theta) = c + (1-c)\frac{\exp[Da(\theta - b)]}{1 + \exp[Da(\theta - b)]}$$

where $D = 1.7$ (the constant that brings the logistic function close to a normal ogive). Then the item information is maximized at the ability

$$\theta_{\max} = b + \frac{1}{Da}\left\{\log[1 + (1+8c)^{1/2}]/2\right\}.$$

In this formula, the symbol "*log*" represents the natural logarithmic function . It may be noted that $\theta_{\max} = b$ when $c = 0$. Hence, for the purpose of ability estimation, Rasch and 2PL items are mapped at their item locations. It may be noted that at $\theta_{\max}$, the probabilities of answering a Rasch or 2PL item either correctly or incorrectly are both equal to 50%.

Type 2: Item Mapping for CR Interpretation and Constructing Ordered Booklets for Bookmark Standard Settings

This process does not deal with ability estimation; rather it is concerned with ways to locate items at various points on a scale so that an expectation can be reasonably attached to that point in terms of what a subject *can do* at that point (Beaton & Allen, 1992). This approach follows closely the writings of Glaser (1963) and Glaser and Nitko (1971) on criterion-referenced (CR) measures and their applications. It assumes the existence of an item pool and has the major function of providing a CR interpretation for selected points (anchor points) on the scale. The creation of an ordered test form for the CTB Bookmark standard setting process (Lewis, Mitzel, & Green, 1996) also falls under this situation of item mapping.

It has been stated (Huynh,1994, 1998) that in the context of CR interpretation and scale anchoring, it may not be appropriate to map the item at the place where the (total) item information is maximized. Let $P(\theta)$ be the probability of answering the item correctly and $Q(\theta) = 1 - P(\theta)$ be the probability of answering the item incorrectly. Then a general form of the Fisher information for a binary item is

$$I(\theta) = [\frac{\delta \log Q(\theta)}{\delta\theta}]^2 Q(\theta) + [\frac{\delta \log P(\theta)}{\delta\theta}]^2 P(\theta).$$

In this formula, the operator δ represents the partial derivative.

It may be noted from the above equation the item information takes into account both probabilities $P(\theta)$ and $Q(\theta)$. Thus, the place where the item information $I(\theta)$ is maximized (the item location) reflects a description for the entire item with both its correct and incorrect responses. Therefore, the item location concept does not embrace any expectation regarding examinee's performance on the item. Huynh (1998, page 36) argues that, in a number of situations, it may be more informative to focus on the location of each separate response. The location of the correct response, for example, might serve as a signal that an examinee located at this place would be "expected" to have the skills underlying the item. This type of item response interpretation appears to be more assertive than a neutral statement that an item is located at a given place. It might therefore be more useful in situations such as attaching CR interpretations to test scores, scale anchoring, and constructing ordered test booklets for Bookmark standard settings. These applications often requires the knowledge of what an examinee can be expected to perform at a given point on the achievement continuum.

Using the Bock partition of item information (Bock, 1972) as a starting point, Huynh (1994, 1998) developed a statistical process to select items for CR and scale anchoring. This procedure results in placing a 3PL item at the ability where the probability of answering the item correctly is

$$p_1 = (2 + c)/3.$$

For a Rasch or 2PL item, this probability is 2/3 or about 67%.

Type 3: Mapping Exemplary Items for Public Release

Here a number of anchor points (such as the Basic, Proficient, and Advanced achievement levels of National Assessment of Educational Progress, NAEP) have already been defined, each with its own subpool of items. The purpose of "item mapping" in this case is to select a number of items that would illustrate the (pre-defined) meaning of that point. The selection of exemplary items to illustrate the three NAEP achievement levels (Bourque, 1997) appears to fall under this aspect of item mapping. For example, to be selected as an exemplary item for the Proficient level of the NAEP 1996 mathematics achievement, the item must meet the following two rules (among other things).

> *Rule 1 (on content): The content of the item must match the content of the operationalized description of the Proficient level.*
>
> *Rule 2 (on probability): The probability of answering the item correctly for a Proficient student must be greater than 51%.* (Bourque, 1997, p. 384)

The first rule on content clearly implies the existence of a set of items appropriate at the Proficient level. This pool is smaller than the total item pool and its size is expected to impact the cutoff probability set in the second rule. For example, if the Proficient item pool is large, one might impose a cutoff probability larger than 51%. On the other hand, if the Proficient pool is small, a smaller cutoff probability might be needed to select the exemplary items. It may be noted that the existence of an item pool at each ability location is not implied in the Type 2 item mapping. In this type of item mapping, an attempt is made to locate a subset of items at a given anchor point without any knowledge about the content of the selected items.

The IRT literature is inundated with well-documented procedures for selecting items for test administration or ability estimation (Type 1 of item mapping). Selection rules for exemplary items (Type 3 of item mapping), by and large, depend on the size of the item pool with content that is appropriate for a given anchor point. Thus it seems fair to state that Type 3 item mapping requires more than a statistical rule for item selection.

Given the statistical nature of this paper, we will now focus only on selection rules for Type 2 item mappings.

Assessing the Consequences of an Item Selection Rule

Within the context of item response theory (IRT), mapping a binary item for CR interpretation is tantamount to replacing the item characteristic function (icf) with a 0/1 step function. Although differing in context, this type of replacement is statistically identical to the process of mastery testing in which test scores (or abilities) are dichotomized into a pass or a failure. A general formulation of mastery testing may be found in Huynh (1976).

Consider now a binary item with increasing (in the large sense) item characteristic function $P(\theta)$. This function represents the probability that an examinee with ability θ will answer the item correctly. Mapping the item at the ability τ_1 for CR interpretation means that an examinee with ability $\theta \geq \tau_1$ is "expected" to answer the item correctly. In the context of mastery testing, the examinee is said to pass the item or is a master. On the other hand, an examinee with ability $\theta < \tau_1$ is "expected" to answer the item incorrectly. In other word, the examinee is deemed as a failure on the item (a non-master) (Huynh, 1998, p. 47). Thus, for CR interpretation, the icf $P(\theta)$ is replaced by the step function

$$s(\theta) = \begin{cases} 0 & if \quad \theta < \tau_1 \\ 1 & if \quad \theta \geq \tau_1. \end{cases}$$

Note that the icf $P(\theta)$ is increasing, so the conditions $\theta \geq \tau_1$and $\theta < \tau_1$ are equivalent to the requirements $P(\theta) \geq p_1$ and $P(\theta) < p_1$ where $p_1 = P(\tau_1)$.

At the ability θ, the probability of answering the item correctly $P(\theta)$ is also the probability of a false negative error. (This is the error encountered when failing an examinee who answers the item correctly.) The probability of answering the item incorrectly $Q(\theta) = 1 - P(\theta)$ represents also the probability of a false positive error. (This error occurs when mastery status is granted to an examinee who answers the item incorrectly.) Now let $C^-(\theta)$ and $C^+(\theta)$ be the cost (or loss) associated with a false negative and false positive error and let $R(\theta)$ be the ratio C^-/C^+. Within a traditional decision-theoretic framework, it is found (Huynh 1976, p. 67) that

$$p_1 = 1/[1 + R]$$

or equivalently

$$R = (1 - p_1)/p_1.$$

Thus, the following relationship holds for the probability p_1 of answering the item correctly at the cutoff ability τ_1.

$$\begin{array}{lll} p_1 < .5 & if & R > 1 \\ p_1 = .5 & if & R = 1 \\ p_1 > .5 & if & R < 1. \end{array}$$

The above formula provides a way to assess the consequences of selecting a value for τ_1 or p_1. For example, if $p_1 = 50\%$, then it may be deduced that $R = 1$ (e.g. the false positive and false negative errors are weighted equally). On the other hand, if $p_1 = 2/3$, then $R = 2$ (e.g. the false negative errors are twice as serious as the false positive errors.)

Empirical Bayes Type 2 Item Mappings: Case of Rasch and 2PL Items

Working within the context of the Bock partition of the item information (Bock, 1972) for Rasch items, Huynh (1994) arrived at the maximum information (MI) selection rule based on the value $p_1 = 2/3$. This cutoff probability also holds for 2PL items. Subsequently, Huynh (1998) developed a general psychometric theory for selecting 3PL items and the categories of polytomous items for CR and scale anchoring. For 3PL items, MI selection rules place an item at the ability place where the probability of answering the item correctly is $p_1 = (2 + c)/3$.

It may be noted that

$$(2 + c)/3 = c + (1 - c) \times 2/3.$$

Thus, the threshold probability of $(2 + c)/3$ for a 3PL item can be deduced from the cutoff probability of 2/3 for a 2PL item (an item without guessing) by using the formula for correction for random guessing.

The remainder of this section provides another way to derive selection rules for Type 2 item mappings. The alternative process is based upon a construction of a "*synthetic*" population of examinees for whom the item is appropriate. In the context of formal mathematical statistics, such a process is similar (if not identical) to the use of an empirical Bayes approach.

To locate a given item on a scale for CR interpretation, the following two questions are posed.

> *Question 1: To which population of examinees is the item most appropriate?*
>
> *Question 2: For the population found in Question 1, what is the typical ability of those who answer the item correctly?*

A formal empirical Bayes solution of Type 2 item mappings within the context of these two questions is provided in Huynh (1998). Starting with a 2PL item and within the family of conjugate and symmetric priors (or ability distributions), it is found that the answer to Question 1 is the ability distribution with a probability density proportional to $[I(\theta)]^\alpha$ where $I(\theta)$ is the item information and α is any positive constant. The corresponding cutoff probability takes the general form

$$p_1 = (\alpha + 1)/(2\alpha + 1)$$

and the cutoff ability is given as

$$\tau_1 = b + \frac{\log[(\alpha + 1)/\alpha]}{Da}.$$

Since α is positive, it is clear from the above formula that, for Rasch and 2PL items, the cutoff probability p_1 is larger than 50%. The maximum information (MI) rule based on $p_1 = 2/3$ satisfies the condition of an empirical Bayes rule.

Empirical Bayes Type 2 Item Mappings: Case of 3PL Items

Consider now a 3PL item with item information $I(\theta)$. As in the case of Rasch or 2PL items, we will consider the ability distribution with probability density proportional to $[I(\theta)]^\alpha$ where α is any positive constant. As documented in Huynh (1998), the cutoff probability p_1^* will now take the general form:

$$p_1^* = \frac{\alpha + c + 1 + [(\alpha + c + 1)^2 - 4c(2\alpha + 1)(1 - \alpha)]^{1/2}}{2(2\alpha + 1)}.$$

As an illustration, let $\alpha = .5$ and $c = .25$. This value for α corresponds to the family of prior (ability) distributions considered by Jeffreys (1938, 1949, 1961) and the value $c = .25$ may be thought as coming from a multiple-choice item with four options. For the situation under study, the cutoff probability is $p_1^* = 79.65\%$. It may be interesting to note that earlier NAEP scale anchoring used a cutoff probability of 80%.

Linking Empirical Bayes Rules for Rasch and 2PL Items to Rules for 3PL Items Through Correction for Random Guessing

For the ability distribution with $\alpha = .5$ of above and with a 2PL item without guessing ($c = 0$), the cutoff probability is $p_1 = (\alpha + 1)/(2\alpha + 1) = 75\%$. If this item had four options (with $c = .25$) and if the formula for correction for random guessing were used, then the cutoff probability would be $p_1^c = .25 + .75 \times .75\% = 81.25\%$. This value differs from the value $p_1^* = 79.65\%$ computed in the previous section. Thus, in general, the cutoff probability p_1^* computed directly from the general empirical Bayes approach for 3PL items differs from the cutoff probability p_1^c computed by applying the formula for correction for random guessing to a 2PL item. *These two cutoff probabilities are identical only when* $\alpha = 1$. When this condition holds, the cutoff probabilities are $p_1 = 2/3$ for Rasch and 2PL items and $p_1^* = (2+c)/3$ for 3PL items. These cutoff probabilities are the ones associated with the maximum information (MI) selection rules derived by Huynh (1994, 1998) based on the Bock (1972) partition of the item information.

Concluding Remarks

Item mappings are widely used in educational assessment. Items are typically mapped in different ways depending on the purpose of the mapping. This paper makes an attempt to categorize item mappings in three broad types with intended use (1) to estimate the ability of an examinee, (2) to attach a criterion-referenced interpretation to test scores and to construct ordered item booklet for Bookmark standard settings, and (2) to select exemplary items for designated anchor points on an achievement continuum. A process is provided for assessing the relative magnitude of the consequences of a given selection rule for item mapping. Rationales and details of an empirical Bayes approach to the construction of selection rules are provided for the family of Rasch, 2PL and 3PL items. It is mentioned that *the empirical Bayes rules are identical to the maximum information (MI) rules when the rules for 3PL items are linked to those of 2PL items though the formula for correction for random guessing.*

References

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics, 17*, 191-204.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika, 37*, 29-51.

Bourque, M.L. (1997). Setting the NAEP achievement levels for the 1996 mathematics assessment. In Allen, N.A., Jenkins, F., Kulick, E., & Zelenak, C.A. (Eds.). *Technical Report (NAEP State Assessment Programs in Mathematics)*. Washington, DC: National Center for Education Statistics.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Glaser, R, & Nitko, A. J. (1971). Measurement in learning and instruction. In R. L. Thorndike (Ed.). *Educational Measurement* (2nd ed., pp. 625-670). Washington, DC: American Council on Education.

Hambleton, R. K., & Swaminathan, H. (1990). *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Huynh, H. (1976). Statistical consideration of mastery scores. *Psychometrika*, 41, 65-78.

Huynh, H. (1994). Some technical aspects of standard setting. In *Proceedings of Joint Conference on Standard Setting for Large-Scale Assessments* (pp. 75-93). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Huynh, H. (1998). On score locations for binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics, 23*, 35-56.

Huynh, H. (September 1998). *On Bayesian score locations for 3PL binary items and their applications to item selection for scale anchoring and criterion-referenced interpretation.* Released at the Response Probability Expert Panel Meeting, National Center for Education Statistics, Washington, DC.

Huynh, H. (April 2000). *On Bayesian rules for selecting binary items for criterion-referenced interpretations and creating ordered item booklets for bookmark standard setting.* Paper presented at the annual meeting of American Educational Research Association, New Orleans.

Jeffreys, H. (1939, 1948, 1961). *The Theory of Probability.* Oxford : Oxford University Press.

Lewis, D.M., Mitzel, H.C., & Green, D.R. (June 1996). *Standard setting: A Bookmark approach.* Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

TM031300

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

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: On item mapping and statistical rules for selecting binary items for criterion-referenced interpretation and bookmark standard setting	
Author(s): HUYNH HUYNH	
Corporate Source:	Publication Date:

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 ☑	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

Signature: Huynh Huynh	Printed Name/Position/Title: HUYNH HUYNH	
Organization/Address: University of South Carolina	Telephone: 803 777 7364	FAX:
	E-Mail Address:	Date: 5/16/2000

HUYNH-HUYNH@SC.ED

(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:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
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:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)