

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 4, Number 6, November, 1995

ISSN=1531-7714

Item Bias Review

Ronald Hambleton and Jane H. Rodgers
University of Massachusetts at Amherst

When important decisions are made based on test scores, it is critical to avoid bias, which may unfairly influence examinees' scores. Bias is the presence of some characteristic of an item that results in differential performance for individuals of the same ability but from different ethnic, sex, cultural, or religious groups.

This article introduces three issues to consider when evaluating items for bias -- fairness, bias, and stereotyping. The issues are presented and sample review questions are posed. A comprehensive item bias review form based on these principles is listed in the references and is available from ERIC/AE. This Article and the review form are intended to help both item writers and reviewers.

In any bias investigation, the first step is to identify the subgroups of interest. Bias reviews and studies generally focus on differential performance for sex, ethnic, cultural, and religious groups. In the discussion below, the term designated subgroups of interest (DSI) is used to avoid repeating a list of possible subgroups.

Fairness vs. Bias

In preparing an item bias review form, each question can be evaluated from two perspectives: Is the item fair? Is the item biased? While the difference may seem trivial, some researchers contend that judges cannot detect bias in an item, but can assess an item's fairness. Perhaps the best approach is to include both types of questions on the review form. (Box 1 offers a list of questions addressing fairness.)

Box 1—Sample Questions Addressing Fairness

- ▶ Does the item give a positive representation of designated subgroups of interest (DSI)?
- ▶ Is the test item material balanced in terms of being equally familiar to every DSI?
- ▶ Are members of DSI highly visible and positively portrayed in a wide range of traditional and nontraditional roles?
- ▶ Are DSI represented at least in proportion to their incidence in the general population?
- ▶ Are DSI referred to in the same way with respect to the use of first names and titles?
- ▶ Is there an equal balance (across items in the test) of proper names? ethnic groups? activities for all groups? roles for both sexes? adult role models (worker, parent)? character development? settings?
- ▶ Is there greater opportunity on the part of members of one group to be acquainted with the vocabulary?
- ▶ Is there greater opportunity on the part of members of one group to experience the situation or become acquainted with the process presented by the items?
- ▶ Are the members of a DSI portrayed as uniformly having certain aptitudes, interests, occupations, or personality traits?

Different Kinds of Bias

Bias comes in many forms. It can be sex, cultural, ethnic, religious, or class bias. An item may be biased if it contains content or language that is differentially familiar to subgroups of examinees, or if the item structure or format is differentially difficult for subgroups of examinees. An example of content bias against girls would be one in which students are asked to compare the weights of several objects, including a football. Since girls are less likely to have handled a football, they might find the item more difficult than boys, even though they have mastered the concept measured by the item (Scheuneman, 1982a).

An item may be language biased if it uses terms that are not commonly used statewide or if it uses terms that have different connotations in different parts of the state. An example of language bias against blacks is found in an item in which students were asked to identify an object that began with the same sound as "hand." While the correct answer was "heart," black students more often chose "car" because, in black slang, a car is referred to as a "hog." The black students had mastered the concept but were selecting the wrong item because of language differences (Scheuneman, 1982b). Questions that might be asked to detect content, language, and item structure and format bias are listed in Box 2.

Box 2- Sample Bias Questions

Content Bias

- ▶ Does the item contain content that is different or unfamiliar to different DSI?
- ▶ Will members of DSI get the item correct or incorrect for the wrong reason?
- ▶ Does the content of the item reflect information and/or skills that may not be expected to be within the educational background of all examinees?

Language Bias

- ▶ Does the item contain words that have different or unfamiliar meanings for DSI?
- ▶ Is the item free of difficult vocabulary?
- ▶ Is the item free of group specific language, vocabulary, or reference pronouns?

Item Structure and Format Bias

- ▶ Are clues included in the item that would facilitate the performance of one group over another?
- ▶ Are there any inadequacies or ambiguities in the test instructions, item stem, keyed response, or distractors?
- ▶ Does the explanation concerning the nature of the task required to successfully complete the item tend to differentially confuse members of DSI?

Stereotyping and Inadequate Representation of Minorities

Stereotyping and inadequate or unfavorable representation of DSI are undesirable properties of tests to which judges should be sensitized. Tests should be free of material that may be offensive, demeaning, or emotionally charged. While the presence of such material may not make the item more difficult for the candidate, it may cause him or her to become "turned off," and result in lowered performance. An example of emotionally charged material would be an item dealing with the high suicide rate among Native Americans. An example of offensive material would be an item that implied the inferiority of a certain group, which would be offensive to that group. Terms that are generally unacceptable in test items include lower class, housewife, Chinaman, colored people, and red man.

Additional terms to avoid include job designations that end in "man." For example, use police officer instead of policeman; firefighter instead of fireman. Other recommendations to eliminate stereotyping:

- Avoid material that is controversial or inflammatory for DSI.
- Avoid material that is demeaning or offensive to members of DSI.
- Avoid depicting members of DSI as having stereotypical occupations (i.e., Chinese launderer) or in stereotypical situations (i.e., boys as creative and successful, girls needing help with problems).

Recommended Reading

This Article is based on Hambleton, R.K. and Rogers, H.J. (1996) *Developing an Item Bias Review Form*, which is available through ERIC/AE.

Berk, R.A. (Ed.). (1982). *Handbook of methods for detecting test bias*. Baltimore, MD: The Johns Hopkins University Press.

Chipman, S.F. (1988, April). *Word problems: Where test bias creeps in*. Paper presented at the meeting of AERA, New Orleans.

Hambleton, R.K., & Jones, R.W. (in press). Comparisons of empirical and judgemental methods for detecting differential item functioning. *Educational Research Quarterly*.

Lawrence, I.M., Curley, W.E., & McHale, F.J. (1988, April). *Differential item functioning of SAT-verbal reading subscore items for male and female examinees*. Paper presented at the meeting of AERA, New Orleans.

Mellenbergh, G.J. (1984, December). *Finding the biasing trait(s)*. Paper presented at the Advanced Study Institute Human Assessment: Advances in Measuring Cognition and Motivation, Athens, Greece.

Mellenbergh, G.J. 1985, April). *Item bias: Dutch research on its definition, detection, and explanation*. Paper presented at the meeting of AERA, Chicago.

Scheuneman, J.D. (1982a). A new look at bias in aptitude tests. In P. Merrifield (Ed.), *New directions for testing and*

measurement: Measuring human abilities, No. 12. San Francisco: Jossey-Bass.

Scheuneman, J.D. (1982b). A posteriori analyses of biased items. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias*. Baltimore, MD: The Johns Hopkins University Press.

Scheuneman, J.D. (1984). A theoretical framework for the exploration of causes and effects of bias in testing. *Educational Psychology*, 19(4), 219-225.

Schmitt, A.P., Curley, W.E., Blaustein, C.A., & Dorans, N.J. (1988, April). *Experimental evaluation of language and interest factors related to differential item functioning for Hispanic examinees on the SAT-verbal*. Paper presented at the meeting of AERA, New Orleans.

Tittle, C.K. (1982). Use of judgmental methods in item bias studies. In R.A. Berk (Ed.), *Handbook of methods for detecting item bias*. Baltimore, MD: The Johns Hopkins University Press.

Descriptors: Cultural Differences; *Culture Fair Tests; Ethnicity; *Evaluation Methods; *Item Bias; Religious Cultural Groups; Sex Differences; *Stereotypes; Test Construction; Test Format; *Test Items

Citation: Hambleton, Ronald & Rodgers, Jane (1995). Item bias review. *Practical Assessment, Research & Evaluation*, 4(6). Available online: <http://PAREonline.net/getvn.asp?v=4&n=6>.