DOCUMENT RESUME

ED 307 333 TM 013 506

AUTHOR Elmore, Patricia B.; And Others
TITLE Counselor Perceptions of Test Bias: Critical Issues in Test Use and Interpretation.
PUB DATE Mar 89
NOTE 20p.; Paper presented at the Annual Meeting of the American Association for Counseling and Development (Boston, MA, March 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Counselor Attitudes; Counselors; Counselor Training; Elementary Secondary Education; Higher Education; *School Counselors; *Test Bias; Testing Problems; Test Interpretation; Test Use; Workshops
IDENTIFIERS *Fairness; Fair Use

ABSTRACT

Thirty-three counselors attending a workshop at the 1988 Annual Convention of the American Association for Counseling and Development were asked to rate six "critical incident" exercises on bias in test content and unfairness in test usage. About two-thirds of the subjects were female, and one-third were male. The subjects ranged in age from 31 to 64 years, with a mean of 45 years. About 30% of the counselors worked in elementary or secondary schools, another 30% were employed in 2-year and 4-year colleges, and the remainder were in a wide variety of other work settings. The six simulation exercises presented to the participants dealt with counselor use of test information at the elementary, secondary, and college levels. Despite special efforts at the beginning of the workshop to define and explain differences between biased test content and unfair test use, many of the participants seemed to be confused and were unable to reach a decision. Few significant relationships between counselors' background and experience and their ratings on these exercises were found. Recentness of having taken a course or workshop in measurement and work setting appeared to be more related to the ratings than did gender, race/ethnicity, age, highest degree, or years of counseling experience. It is recommended that workshops or inservice programs be developed for counselors to instruct them in means of detecting test bias and using tests fairly. (JH)

Counselor Perceptions of Test Bias: Critical Issues in Test Use and Interpretation

Patricia B. Elmore

Southern Illinois University-Carbondale

Esther E. Diamond

Educational and Psychological Consultant

Evanston, Illinois

Ruth B. Ekstrom

Educational Testing Service

Paper presented at the Annual Meeting of the

American Association for Counseling and Development

Boston

March, 1989

Abstract

Thirty-three counselors attending workshop were asked to rate six "critical incident" exercises on bias in test content and unfairness in test usage. Despite spending time at the beginning of the workshop to define and explain the difference between biased test content and unfair test use, many of the participants appeared to be confused and unable to reach a decision. Few significant relationships between counselors' background and experience and their ratings on these exercises were found. Recentness of having taken a course or workshop in measurement and work setting appeared to have more relationship to the ratings than did gender, race/ethnicity, age, highest degree, or years of counseling experience. It is recommended that workshops or in-service programs be developed for counselors to learn about ways of detecting bias and using tests fairly.

Counselor Perceptions of Test Bias: Critical Issues in Test Use and Interpretation

The National Commission on Excellence in Education (1983) described tests as important tools in a student's education, tools for identifying strengths and weaknesses and pinpointing needs for educational remediation. A report by the Institute on Research for Teaching (1980) pointed out that testing, for the most part, should be used to monitor student progress rather than to rate it. In working with clients, counselors often make use of achievement tests, aptitude tests, and interest inventories. Despite the important role that counselors play in test selection and use, little is known about counselors' ability to detect test bias and to interpret and use tests fairly and equitably.

The report of the Institute on Research for Teaching (1980) was particularly critical of the inadequate training of educational personnel in test use and the understanding of various kinds of test scores. Schafer & Lissitz (1987) reported that a significant proportion of school personnel receive little training in measurement, although school counselors receive more than others.

Teachers, counselors, and others who select or use tests need to know what to look for in a test, how to select tests appropriate for particular purposes, whether the tests were appropriately developed, and how to interpret scores and norms correctly. They need to determine whether the tests they are considering are sensitive to individual and group differences in culture, experiences, and other important factors. They need to know whether the inferences to be drawn from the results are supportable, based not only on the test score but also on other relevant information about the student. Finally, they need to determine whether the test is really necessary--whether the results would add new dimensions to information already available, or merely confirm what is already known about the student.

These responsibilities of test users--alongside those of test developers--are set forth in detail in the Code of Fair Testing Practices in Education (1988), endorsed by AACD and AMECD. Separate sections deal with developing/selecting appropriate tests, interpreting scores, striving for fairness, and informing test takers about the purpose and content of the test, the types of items it contains, appropriate test-taking strategies and the like, as well as students' rights as test takers.

Bias in testing has been a concern of the Association for Measurement and Evaluation in Counseling and Development (AMECD) (formerly the Association for Measurement and Evaluation in Guidance [AMEG]) since 1970. Two articles published by the AMEG Commission on Sex Bias in Measurement (1973, 1977) reported on sex bias in interest inventories. A subsequent study (Diamond, 1980) surveyed publishers of standardized achievement tests to determine the techniques used to minimize sex bias. In 1983 the AMECD Committee on Bias in Measurement expanded its charge to include an investigation of not only sex bias but also ethnic and minority bias in standardized achievement tests. Diamond and Elmore (1986) again surveyed publishers of standardized achievement tests and, although improvement over the earlier results was found, concluded that the effort to detect and minimize test bias must be a dynamic, on-going process and the joint responsibility of all stakeholders, including those who develop the tests, those who publish them, those who select them and use the results, and those who are affected by the way in which the results are used.

As part of the activities of the AMECD Committee on Bias in Measurement, the authors of this paper (Elmore, Diamond, & Ekstrom, 1988) asked a group of counselors to indicate whether or not they thought test bias, as

distinguished from unfairness, occurred in six situations. The counselors were also asked to indicate what might explain the findings presented in each situation and how a counselor might respond to the situation. The results provide an initial insight into how counselors perceive test bias and unfairness. The information collected in this preliminary survey is a first step in understanding how counselors use tests and, through this information, in providing counselors with education that will further appropriate test use.

Method

Subjects

The subjects were 33 counselors attending a workshop at the 1988 American Association for Counseling and Development (AACD) Annual Convention in Chicago. Approximately two-thirds (66 percent) of the subjects were females and one-third were males. Seventy-six percent of the subjects were white, 21 percent Black, and three percent Hispanic. The subjects ranged in age from 31 to 64, with a mean age of 45. The majority of the counselors (63 percent) had received a master's degree; 18 percent reported a specialist degree and another 18 percent a doctorate. The number of years of counseling experience ranged from one to 30, with a mean of 12 years of experience. The mean number of years

since the participants took a course or workshop in measurement was 5 and the range was from 0 to 25 years. About 30 percent of the counselors worked in elementary or secondary schools, another 30 percent were employed in 2-year and 4-year colleges, and the remainder were in a wide variety of other work settings including industry, employment counseling centers, and private practice.

The counselors were asked to name the tests which they used most often. The most frequently named tests were the Strong-Campbell Interest Inventory, the Myers-Briggs Type Indicator, the American College Testing Program, and the Scholastic Aptitude Test. A total of 62 different tests and inventories were named.

Instrument

The six simulation exercises presented to the participants dealt with counselor use of test information at the elementary, secondary and college level. The content was based on textbook discussions of bias and experiences of the committee members. Exercises 1 and 2 were adapted from a chapter by L. A. Shepard (1982, pp. 11-12) in R. A. Berk's Handbook of Methods for Detecting Test Bias. The text of the six exercises follows:

1. An elementary school counselor notices that on an achievement test there is a large difference between the scores of Black and White students on the Word Problems subtest of the Mathematics

portion of the test. The counselor gives these same test problems orally to a group of students and finds no difference between the proportion of White and Black students solving the problems correctly.

2. A high school counselor is asked by the principal to give a verbal analogies test to two groups of students to determine their reasoning ability. Both groups of students came to the United States three years ago and now read English about equally well. One group is from Japan and one is from Italy. The counselor reviews the content of the test and notes that 80 percent of the items are based on words with Latin origins.

3. A newly appointed college counselor notices that twice as many males as females receive a special state scholarship for outstanding students. When she inquires, she finds that the students are selected by adding the Verbal and Mathematics scores on the college's admission test, and then taking the students with the highest combined scores.

4. A junior high counselor is confronted by a distraught seventh grade student. The student was selected to participate in the statewide Talent Search during sixth grade because she had regularly scored at the 95th percentile on standardized achievement tests during elementary school. As part of the Talent Search the student was administered a college admission test usually given to high school juniors and seniors. The student scored below the 25th percentile compared to other Talent Search students on the college admission test and feels she is a failure.

5. Fran, a college freshman, has taken an interest inventory that compares her responses with those of both men and women in a variety of different occupations. However, some of the occupations have been normed only on men and some only on women. In general, Fran's highest scores are those on women's norms, for occupations such as teacher, librarian, and social worker--occupations in which women have generally predominated. Her highest scores on male norms, however, are for occupations such as personnel manager, industrial psychologist, and

architect, for which female norms have not yet been developed.

6. Lisa has just taken an interest inventory that measures vocational interests in a number of broad areas. Lisa's responses are compared with those of girls at her grade level (ninth). Her score on the Scientific scale is at the 70th percentile. On the Mechanical scale it is at the 73rd percentile. Her highest score is on the Clerical scale. The counselor comments that boys with the same raw scores on the Scientific and Mechanical scales would score at much lower percentiles, compared with boys, but on the Clerical scale their scores would be at a higher percentile.

The workshop participants were asked to indicate (by checking their preferred response for each item) if they considered the test/inventory to be biased or unbiased and the use to be fair or unfair. The subjects could respond don't know to either question. The participants were also asked to provide a short, open-ended response indicating what might explain the situation and how a counselor might respond to it.

Procedure

Before the subjects were presented with the six simulation exercises, a brief overview of the difference between test bias and unfairness in testing was presented. The following definitions from Diamond and Tittle (1985) were distributed to the participants:

Bias: refers to the intrinsic characteristics of a test; that is, to the "content, the construct or constructs the test is supposedly measuring, and the context within which the content is embedded" (p. 168). Bias is considered to occur "when two individuals of equal ability but from different groups respond differently to a test item and therefore do not have the same probability of success on the item" (p. 168).

Unfairness: refers to "ethical questions involving use of the test results" (p. 168).

Instructions to participants empnasized the importance, when members of different groups exhibit differences in test scores, of determining whether or not the members of each group have had equal opportunity to learn the material which the test is assessing. If tests are reflecting such differences in opportunity to learn, group score differences are not necessarily indications of test bias but are more likely indicators of inequity or other problems in the educational or counseling system.

A discussion session followed the presentation of each exercise. In each discussion there was a show of hands to indicate the bias and fairness responses. There was also an opportunity for participants to indicate their responses

regarding the causes of the score differences and possible counselor responses These discussions may have, in some cases, led some subjects to modify or add to their original responses. This may make these responses less valid than if they had been obtained in a situation with no feedback but, given the workshop setting, a more test-like data collection seemed inappropriate.

Results

The first part of the analysis focused on the participants' perceptions of bias and fairness. The percentages of workshop participants responding in each category for bias (biased, don't know, and unbiased) and for unfairness (unfair, don't know, and fair) for each of the six exercises are shown in Table 1. The test in Exercise 2

- - - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - - -

was perceived as biased by the largest group of subjects while the test use situations in Exercises 3 and 4 were most often perceived as unfair. However, substantial numbers of counselors were not able to evaluate test bias and unfairness in these exercises. The percentage of counselors choosing the don't know response option was as high as 45% and 48% in the bias and unfairness ratings, respectively.

The highest don't know rating for test bias was Exercise 4 and for unfair test use was Exercise 5. Mean ratings for test bias and unfair usage for each exercise are also shown in Table 1. When considering mean ratings, the workshop participants rated Exercise 2 as the most biased test use and Exercises 4, 3, and 2 the most unfair test use.

In the second part of the analysis the perceptions of test bias and unfair test usage were related to the background characteristics and experiences of the workshop participants.

Correlations between the exercise ratings and the three continuous variables (age, years of counseling experience, and time since last workshop or course in measurement) were computed. The statistical hypothesis that the population correlation coefficient was not different from zero was tested. Age and the number of years of counseling experience were not significantly related to the ratings of test bias or unfair test use for any of the exercises. Time since the last measurement course or workshop was significantly related to ratings of unfairness on Exercise 2 ($r = .46$, $p < .05$) and ratings of bias on Exercise 3 $r = -.47$, $p < .05$) but not to the other ratings.

To determine if counselors responded differently to the exercises on the basis of gender and race/ethnicity (white,

minority), t-tests were conducted. No differences were found between male and female counselors' ratings of test bias on Exercises 1, 2, 3, 4, and 5 and of unfair test use on all six exercises. No differences were found between white and minority counselors' ratings of test bias on Exercises 1, 3, 4, 5, and 6 and of unfair test use on Exercises 3, 5, and 6. Unfortunately, lack of variance within responses for one of the groups made it impossible to make the gender comparisons for ratings of test bias on Exercise 6 and the race/ethnicity comparisons for ratings of test bias on Exercise 2 and test usage on Exercises 1, 2, and 4.

One-way analyses of variance were conducted to determine if bias and unfairness responses to the six exercises differed according to the highest degree attained (masters, specialist, doctorate) or work setting (elementary and secondary schools, community colleges and universities, and other settings) of participants. There were no significant differences by highest degree attained. Differences in ratings of test bias were found by work setting on Exercise 4, F (2, 21) = 4.38, $p < .05$. The Scheffe' multiple comparison procedure indicated that the mean test bias ratings of elementary/secondary school counselors (M = 2.00) was significantly different than the

mean test bias ratings of community college/university counselors (M = 2.83) but the ratings of these two groups did not differ significantly from ratings of counselors in non-educational settings (M = 2.36). Differences in ratings of unfair test use were found by work setting on Exercise 5, F (2, 18) = 3.63, $p < .05$. The Scheffe' multiple comparison procedure indicated that no pairwise comparisons were significant.

Discussion and Conclusions

Thirty-three counselors attending a workshop at the 1988 AACD Annual Convention were asked to rate six "critical incident" exercises to indicate whether or not there was indication of: (a) bias in test content, and (b) unfairness in test usage. Despite spending time at the beginning of the workshop to define and explain the difference between biased test content and unfair test use, many of the participants appeared to be confused and unable to reach a decision.

The counselors rated the clearest evidence of bias in test content as occurring in Exercise 2, which describes a verbal reasoning test in which a large proportion of the words have Latin cognates. The counselors' ratings indicate that they believed the college admission tests described in

Exercises 3 and 4 and the interest inventory described in Exercise 6 were probably unbiased in their content.

The counselors gave the highest ratings for unfair test use to the incidents in Exercises 2, 3, and 4. In Exercise 2, two groups of students whose native languages are not English are being asked to take a verbal reasoning test which has many words related to the native language of one group but not the other. In Exercise 3, a college is using admission test scores, in isolation from other information about the students, to award scholarships. In Exercise 4, a college admission test containing content to which they have not been exposed is being used to screen junior high school students for a Talent Search program. The counselors gave a fair test use rating to Exercise 6, which describes a counselor telling a student about the different male-female percentile rankings on an interest inventory. The highest percentage of don't know ratings for unfairness for Exercise 5 indicates that counselors may be unfamiliar with the problem of gender norms in interest measurement and need to know how differences in socialization influence inventory scores.

There appeared to be few significant relationships between counselors' backgrounds and experiences and their ratings on these exercises. Recentness of having taken a

course or workshop in measurement and work setting appeared to have more relationship to the ratings than did gender, race/ethnicity, age, highest degree, or years of counseling experience. This suggests that counselors who have learned about the problems of test bias and test misuse, either through formal study or informal learning in their place of employment, are better prepared to deal with these problems than counselors who have not received instruction on these topics. Therefore, we recommend that AACD present more workshops for counselors to learn about ways of detecting bias and using tests fairly. We also recommend that AACD members prepare journal articles and other materials on this topic that can be used in self-study or in-service programs for counselors who are unable to attend the annual meeting.

References

AMEG Commission on Sex Bias in Measurement. (1973). AMEG Commission report on sex bias in interest measurement. Measurement and Evaluation in Guidance, 6, 171-177.

AMEG Commission on Sex Bias in Measurement. (1977). A case history of change: A review of response to the challenge of sex bias in career interest inventories. Measurement and Evaluation in Guidance, 10, 148-152.

Code of Fair Testing Practices in Education. (1988). Washington, DC: Joint Committee on Testing Practices.

Diamond, E. E. (1980). The AMEG Commission Report on sex bias in achievement testing. Measurement and Evaluation in Guidance, 13, 135-147.

Diamond, E. E., & Elmore, P. B. (1986). Bias in achievement testing: Follow-up report of the AMECD Commission on Bias in Measurement. Measurement and Evaluation in Counseling and Development, 19, 102-112.

Diamond, E. E., & Tittle, C. K. (1985). Sex equity in testing. In S. Klein (Ed.), Handbook for achieving sex equity through education (pp. 163-188). Baltimore, MD: The Johns Hopkins University Press.

Elmore, P. B., Diamond, E. E., & Ekstrom, R. B. (1988, March). Using tests responsibly in counseling. Paper

presented at the meeting of the American Association for Counseling and Development, Chicago.

Institute for Research of Teaching. (1980). Integrating assessment with instruction: A review. (Research series No. 75). Michigan State University: Author.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: U. S. Government Printing Office.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38 (3), 57-63.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 9-30). Baltimore, MD: The Johns Hopkins University Press.

Table 1

Percentage Responding by Category and Mean Ratings of Bias and Unfairness

Exercise	Percentage			Mean[a]	Percentage			Mean[b]
	Biased	Don't Know	Unbiased		Unfair	Don't Know	Fair	
1	50	20	30	1.80	50	39	11	1.61
2	82	4	14	1.32	75	21	4	1.29
3	21	21	57	2.36	85	4	11	1.26
4	10	45	45	2.34	86	4	11	1.25
5	32	39	29	1.96	28	48	24	1.96
6	0	17	83	2.83	24	29	48	2.24

[a]1 = Biased, 2 = Don't Know, 3 = Unbiased

[b]1 = Unfair, 2 = Don't Know, 3 = Fair