

The Sex Fairness of Unnormed Interest Inventories

LINDA S. GOTTFREDSON

Prediger's recent (1981) article is one of a series of published articles in which Prediger and his colleagues at the American College Testing Program (ACT) have questioned the usefulness and fairness of the (unnormed) Self-Directed Search (SDS) for females. Since his article is part of a systematic effort to evaluate the SDS specifically and unnormed interest tests in general, I will first address some problems with the article itself and then discuss the larger issue of sex differences and same-sex test norming (as used, for example, in the inventory competing with the SDS published by Prediger's own organization, ACT).

THE LOGIC OF ANALYSIS

Prediger argued that "construct validity appears to have more relevance than predictive validity for common counseling uses of the SDS and other measures of interests" (p. 118). He then showed that, according to his definition of construct validity, sex-normed SDS scores provide "more accurate estimates of [women's] true interests" (p. 127) than do raw or unnormed scores. I will illustrate the difference between what Prediger referred to as construct validity and predictive validity by presenting data from his Table 1 in a different form. This illustration shows that Prediger's claim of greater accuracy of normed scores is misleading. Since Prediger performed the same type of re-analysis on all of the studies he listed in his Table 2, as Table 1 stands or falls, so do all of his other re-analyses.

Prediger's Table 1 was based on data from a study by G. Gottfredson and Holland (1975), in which they assessed college women's vocational interests with the SDS (using both normed and unnormed scores) and then asked for the women's occupational aspirations (expressed choices) one or three years later. The upper half of my Table 1 shows the Gottfredson-Holland data along with the results of Prediger's re-analysis as published in Prediger's Table 1. The lower half of my table shows some

Linda S. Gottfredson is a research scientist at the Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Maryland.

TABLE 1
Calculations Illustrating Prediger's and an Alternative Method of Assessing the Validity of Raw versus Normed SDS Scores

Data Presented in Prediger's Table 1												
Criterion Group (according to expressed occupational choice)												
SDS Scale	Raw Scores						Normed Scores					
	R	I	A	S	E	C	R	I	A	S	E	C
R	0	2	0	4	0	1	3	16	15	61	4	2
I	3	63	7	68	8	4	1	72	5	100	6	4
A	0	21	57	81	11	3	0	19	54	116	11	3
S	1	50	46	481	24	8	0	11	13	155	7	1
E	0	3	5	9	2	1	0	12	25	124	10	1
C	0	0	2	14	3	7	0	9	5	101	10	13
Predominant interest for group	(I)	I	A	S	(S)	(S)	R	I	A	S	(A)	C

Calculations from the Data Above												
SDS Scale	"Construct validity": Group hits. (Is the most frequent interest type congruent with the criterion group?)			"Predictive validity": Individual hits. (The proportion of women in each criterion group whose interests and choices are congruent.)			Percentage of Women in Each Type ^a					
	Raw	Normed	Expressed Choices	Raw	Normed	Expressed Choices	Raw	Normed	Expressed Choices			
R	No	Yes	0/4	3/4	0.7	10.2	0.4					
I	Yes	Yes	63/139	72/139	15.3	19.0	14.1					
A	Yes	Yes	57/117	54/117	17.5	20.5	11.8					
S	Yes	Yes	481/657	155/657	61.7	18.9	66.4					
E	No	No	2/48	10/48	2.0	17.4	4.9					
C	No	Yes	7/24	13/24	2.6	14.0	2.4					
Total	50%	83%	62%	31%								

^aThese are simply the marginal percentages for the data in the upper half of the table.

calculations on those data that reveal Prediger's special logic and an alternative way of summarizing the data. These calculations provide the basis for the discussion that follows.

The first two columns in the lower half of the table summarize Prediger's analysis of "construct validity." If we ask whether the most frequent SDS interest is congruent with the expressed aspirations of the group, we indeed see that the answer is "yes" for 3 of the 6 groups with raw scores and for 5 of 6 with normed scores. For example, of the women with investigative occupational choices, most also had previously scored highest on the investigative scale of the SDS; this was true for both the unnormed and normed SDS scores.

But what happens if we look at degree of interest-aspiration congruence for the 989 individual women rather than for the 6 criterion groups (the middle columns in Table 1)? In this instance we see that the raw scores are superior, with an overall total of 62% hits compared to only 31% for normed scores. We also see that the two expressed choice criterion groups that account for the greater "construct validity" of normed scores (R and C) are the least popular of the six groups, and together account for only 3% of the sample. In contrast, the most popular group (S), which accounts for 66.4% of the sample, is considered equally construct valid according to Prediger's criterion even though only one-third as many of these women appear to have the "correct" interests for their occupational aspirations when assessed according to normed rather than raw scores.

Which results should we rely on for counseling purposes? Using the same initial data, one analysis clearly favors normed scores and one favors raw scores. Prediger argued that construct validity is more useful in this context than predictive validity, but it is also true that predictive validity is often treated as one indication of construct validity (Cronbach & Meehl, 1955). The real difference is between group-wise and individual rates, not between predictive and construct validity. Merely labelling his group-wise re-analyses of the same longitudinal data "construct validity" does not transform Prediger's work into something other than another type of predictive validity study. If Prediger objects to the use of predictive validity studies, then a study with new data and a different criterion (perhaps occupational satisfaction rather than later choices) would be more convincing than a redefinition of validity.

The last three columns in Table 1 provide an even more compelling reason to question the logic of Prediger's re-analyses. The columns show the percentages of women classified into the six types by their raw and normed SDS scores and by their expressed aspirations, results that Prediger did not mention. The distribution of raw scores is quite similar to that of the expressed choices, with over 60% of women falling into the S category but fewer than 1% into the R category in both cases. The distribution of normed scores, however, diverges sharply from the distribution of expressed choices because normed scores are more evenly distributed across the types—another advantage of normed scores according to Prediger. But if Prediger is going to take expressed choices

calculations on those data that reveal Prediger's special logic and an alternative way of summarizing the data. These calculations provide the basis for the discussion that follows.

The first two columns in the lower half of the table summarize Prediger's analysis of "construct validity." If we ask whether the most frequent SDS interest is congruent with the expressed aspirations of the group, we indeed see that the answer is "yes" for 3 of the 6 groups with raw scores and for 5 of 6 with normed scores. For example, of the women with investigative occupational choices, most also had previously scored highest on the investigative scale of the SDS; this was true for both the unnormed and normed SDS scores.

But what happens if we look at degree of interest-aspiration congruence for the 989 individual women rather than for the 6 criterion groups (the middle columns in Table 1)? In this instance we see that the raw scores are superior, with an overall total of 62% hits compared to only 31% for normed scores. We also see that the two expressed choice criterion groups that account for the greater "construct validity" of normed scores (R and C) are the least popular of the six groups, and together account for only 3% of the sample. In contrast, the most popular group (S), which accounts for 66.4% of the sample, is considered equally construct valid according to Prediger's criterion even though only one-third as many of these women appear to have the "correct" interests for their occupational aspirations when assessed according to normed rather than raw scores.

Which results should we rely on for counseling purposes? Using the same initial data, one analysis clearly favors normed scores and one favors raw scores. Prediger argued that construct validity is more useful in this context than predictive validity, but it is also true that predictive validity is often treated as one indication of construct validity (Cronbach & Meehl, 1955). The real difference is between group-wise and individual rates, not between predictive and construct validity. Merely labelling his group-wise re-analyses of the same longitudinal data "construct validity" does not transform Prediger's work into something other than another type of predictive validity study. If Prediger objects to the use of predictive validity studies, then a study with new data and a different criterion (perhaps occupational satisfaction rather than later choices) would be more convincing than a redefinition of validity.

The last three columns in Table 1 provide an even more compelling reason to question the logic of Prediger's re-analyses. The columns show the percentages of women classified into the six types by their raw and normed SDS scores and by their expressed aspirations, results that Prediger did not mention. The distribution of raw scores is quite similar to that of the expressed choices, with over 60% of women falling into the S category but fewer than 1% into the R category in both cases. The distribution of normed scores, however, diverges sharply from the distribution of expressed choices because normed scores are more evenly distributed across the types—another advantage of normed scores according to Prediger. But if Prediger is going to take expressed choices

Personally, I want my daughters to receive inventory results that reflect what they say their interests are relative to all people, not just relative to other women. This means using raw and not normed scores. If their interests turn out to be stereotypically feminine, *then* they can be encouraged in other ways to examine whether or not those interests are really their own or are primarily reflections of what they think is expected of them as women. The goal of vocational assessment should be to help youngsters of both sexes find jobs they like and are competent in; it should not be to get them into jobs that conform to someone else's vision of what society should look like.

REFERENCES

- Cronbach, L.J., & Meehl, P.E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, 52, 281-302.
- Gottfredson, G.D., & Holland, J.L. Vocational choices of men and women: A comparison of predictors from the Self-Directed Search. *Journal of Counseling Psychology*, 1975, 22, 28-34.
- Gottfredson, L.S. Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology Monograph*, 1981, 28, 545-579.
- Herzog, A.R. High school seniors' occupational plans and values: Trends in sex differences 1976 through 1980. *Sociology of Education*, 1982, 55, 1-13.
- Prediger, D.J. A note on Self-Directed Search validity for females. *Vocational Guidance Quarterly*, 1981, 30, 117-129.