

Alternatives for Validating Interest Inventories Against Group Membership Criteria

Dale J. Prediger

The American College Testing Program

Interest inventories are frequently validated against group membership criteria. Two approaches are considered, only one of which is commonly used. The choice between the two approaches depends on the application being validated. The first and most common approach assumes that interest inventories are to be used in predicting which occupation counselees will enter or prefer. The second assumes that interest inventories are to be used in suggesting occupations for counselees to consider on the basis of compatibility of interests. Validation of these two uses of interest inventories requires different treatment of criterion group base rates. As illustrated by data drawn from a published study, the two approaches to validation can produce substantial differences in criterion group hit rates. Such differences may be found in any study validating group membership predictions if criterion group sizes vary greatly.

The judgment of a person's similarity to members of various occupational groups has long been the cornerstone for counseling uses of interest inventories. Inventories such as Kuder Occupational Interest Survey and Strong Vocational

Interest Blank have traditionally reported results in terms of similarity of a person's measured interests to occupational groups. As Kuder (1970) has noted, "the problem of establishing validity for counseling purposes becomes one of classification [hence] one of the fundamental questions in judging a vocational interest inventory is how well it differentiates among the specific occupational groups for which it is scored" (p. 209). Strong (1943), although primarily concerned with the differentiation of occupational groups from men or women in general, also recognized the need to differentiate among the occupational groups themselves. Though other, perhaps better, approaches to validation are possible (e.g., determination of correlations with satisfaction or success), interest inventory construction and validation studies have typically focused on criterion group differentiation/classification.

In the current literature, the criterion-related validity of an interest inventory is frequently reported in terms of its ability to accurately identify members of various occupations or occupational preference groups (e.g., see Borgen, 1972; Campbell, 1971; Dolliver, 1975; Gottfredson & Holland, 1975; and Zytowski, 1976). The validation process essentially involves the comparison of group membership predictions based on interest scores with actual group membership.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 2 Spring 1977 pp. 275-280
© Copyright 1977 West Publishing Co.

(For purposes of this discussion, the specific prediction procedure—e.g., high-point code, clinical intuition, maximum-likelihood classification incorporating antecedent probabilities—is not directly relevant.) Once group membership predictions have been made, the number of correct predictions are tallied for each of the criterion groups. At this stage of the validation process, researchers face a crucial decision about how to treat variations in the size (base rates) of the criterion groups.

In previous discussions of the validity of group membership predictions (e.g., see Meehl & Rosen, 1955), it is usually assumed that the goal of assessment is to maximize overall “hit rate” (i.e., the overall accuracy of the predictions). This has been the traditional goal of assessment in institutional applications of tests, for example, in selection and placement. In some applications of tests to personal counseling, however, the goal of assessment may be to maximize the hit rate within each of the criterion groups. The manner in which criterion group base rates are used in validity analyses reflecting these two goals can produce substantial differences in hit rates obtained for the same study. This paper examines two alternatives to the use of base rates in validating psychological assessments against group membership criteria. The application and consequences of each alternative are illustrated by examples drawn from the field of interest measurement.

Rationale for Validation Alternatives

It is well known that the validity of a measuring instrument depends on the purposes for which it is used. Hence, before studying validity, one must ask, “Validity for what?” Interest inventories are commonly used to suggest possible occupational options to counselees. Yet, the validity of inventories is often reported in terms of their ability to predict future occupational preferences or occupational entry. As Berdie (1970) has noted, few counselors are interested in predicting whether a counselee will enter (or prefer) occupation A or occupation B. Hence, validity data for this use of interest inventories

may provide a distorted view of validity for more common uses. Some of the reasons are discussed below.

The “Will-Prefer-or-Enter” Criterion

When predicting the occupations persons will prefer or enter, the nature of employment distributions as well as the nature of occupational preferences must be taken into account. Stated another way, if an interest inventory is to provide accurate predictions of eventual employment, the predictions must accurately reflect the size of each occupational criterion group. To the degree that group membership predictions depart from group base rates, the inventory’s predictive accuracy will be lowered.

Interest inventories predicting that persons will enter or prefer occupations in the same proportions as in the past should do well under this approach to validation. For a multitude of reasons (e.g., social expectations, local labor market needs, the contingencies of life), people will continue to state preferences for and enter traditional occupations. Unfortunately, the number of persons in various occupations and occupational preference groups differs widely from group to group (Gottfredson, Holland, & Gottfredson, 1975; Prediger, Roth, & Noeth, 1974). Since the predictions used in validation studies are based on the same scores counselees receive, the occupational options suggested to counselees will reflect the same differences in base rates as the predictions. Under this approach to validation, a “valid” interest inventory in the 1850s would have suggested farming to nearly everyone. The employment status quo will be reflected and reinforced by interest inventories validated in this way.

The “Should-Consider” Criterion

The alternative approach to the use of occupational preference and membership as criteria in validating interest inventories assumes that the purpose of interest inventories is to identify career options for counselees to consider rather than to predict the occupations counselees will prefer or enter. To achieve the former objective,

an interest inventory must assess the correspondence between a counselee's interests and the interests associated with various occupational groups—regardless of the group base rates. If a counselee's interests are similar to the interests of persons in a given occupation, one would suggest that the counselee consider the occupation, even though relatively few persons are employed in the occupation. The emphasis is on "should consider," not "will enter or prefer." The underlying assumption is that employment data may play an important role in career counseling, but they should not influence interest score reports.

Studies following this approach to interest inventory validation will treat occupational criterion groups (or preference groups) as if they were of equal size. One would expect an interest inventory to suggest engineering to a large proportion of criterion group members in engineering, nursing to a large proportion of nurses, retail sales to retail sales clerks, horseshoeing to horseshoers, and so on for each of the criterion groups available. The fact that there are relatively few horseshoers in comparison to retail sales clerks is irrelevant. The question asked in this validation analysis is "What proportion of the members of *each* criterion group would have been asked to look into their occupation by this interest inventory?" Stated differently, the question is "What is the hit rate for each criterion group?" A high hit rate depends on an inventory's ability to differentiate the criterion groups and, thus, minimize the misassignment of members of each of the groups.

In this approach to validation, an interest inventory does not have to suggest retail sales to more counselees than horseshoeing because there are more retail sales clerks than horseshoers. "Predictions" are simply based on whichever criterion group a person resembles most. There is no premium placed on providing interest-score distributions that parallel preference or employment distributions. This proposed validation strategy recognizes that, for a number of very practical reasons, many persons may not enter the occupations suggested ("predicted") by an interest inventory.

How Choice of Criterion Affects Career Guidance

Perhaps the following example will bring differences between the two approaches to validation into sharper focus. Suppose that in a society built on the caste system, an interest inventory was designed to have high validity in predicting occupational entry. The inventory would suggest few, if any, occupations that were not traditional for a person's caste. To do otherwise would lower its validity. On the other hand, suppose the inventory was designed to identify occupational options compatible with a person's interests—regardless of the proscriptions of the society. Such an inventory may suggest many occupations not traditional for members of the caste. As a result, it would be a poor predictor of occupational entry. Yet, it may do an excellent job of determining occupational compatibility. Even in a time of social change, the score reports might be unsettling, but they could provide beneficial information, both to the individual and to the society. Prediger and Cole (1975) provide an extended discussion of this topic, emphasizing the use of interest inventories in counseling males and females concerning nontraditional careers.

Implications for Validation Procedures

It should be apparent from the above discussion that the essential difference in the two approaches to interest-inventory validation lies in the way in which criterion group base rates are treated. Two basic options are described below.

Option 1: In determining predictive accuracy, use the weighted average hit rate. This option is often chosen, perhaps by default. When criterion group size is "ignored" (i.e., when "hits" are simply totaled across the groups), the relative sizes of the criterion group samples determine the weighting. Hence, when Option 1 is followed, the overall hit rate is a *weighted average hit rate*. That is, the hit rates for each criterion group are weighted according to the criterion group base rates. Option 1 is preferable when trying to predict occupational entry. As noted later, however, other bases for differential

weighting might be more appropriate when this criterion is chosen.

Option 2: In determining predictive accuracy, use the unweighted average hit rate. This option gives each of the criterion groups equal importance in determining predictive accuracy. The objective, in effect, is to maximize the level of predictive accuracy within *each* of the criterion groups. Hence, the overall hit rate is the *unweighted* average hit rate for the criterion groups. The hit rate for a large criterion group cannot overwhelm the hit rates for several smaller groups, as in Option 1. Attention is drawn to predictive effectiveness within each of the criterion groups. Option 2 would appear to be appropriate for validating an inventory designed to suggest occupations that counselees might want to consider.

Illustration of the Two Approaches to Validation

Application of the two options to recent research results will show that the consequences of their use are far from academic. In a study comparing predictions of occupational preference obtained from Self-Directed Search (SDS), raw scores and standard scores based on same-sex norms—Gottfredson and Holland (1975)—conclude that when standard scores were used with college women, “predictive validity decreased greatly” (p. 32). Criterion group hit rates in the Gottfredson-Holland study are summarized in Table 1. Holland’s occupational typology is used to identify the criterion groups, and data for the two samples of college women in the study are combined to increase criterion group size. Predictions for both raw scores and standard scores are based on the high-point code (i.e., highest score) for the six scales in Holland’s SDS.

Results and Implications

Table 1 shows that when Option 1 is chosen, SDS raw scores produce a weighted average hit rate that is twice as large as the hit rate for standard scores. Hence, results for Option 1 support the use of raw-score reports of interests, the usual procedure for the SDS (Holland, 1972). If

Option 2 is chosen, however, the overall hit rates for SDS raw scores and standard scores are essentially the same. On the basis of the equivalent hit rates obtained under Option 2, one might decide that other factors should be given primary importance in reaching a decision concerning the use of raw-score or standard-score reporting procedures. This decision could be of considerable consequence. For example, it has been shown that widely divergent career options are suggested to males and females by raw scores on the SDS and other interest inventories (Holland, 1972; Gottfredson, Holland, & Gottfredson, 1975; Prediger & Hanson, 1974). In contrast, standard scores (based on same-sex norms) suggest similar career options (Gottfredson et al., 1975; Prediger & Hanson, 1974). In addition, psychometric theory favors the use of norms in the comparison of assessments of human traits.

Discussion

In this study, the difference in results obtained for Options 1 and 2 is primarily due to the large size of the Social Group. The high raw-score hit rate for this group, in combination with size, makes a major contribution to the hit rate obtained under Option 1. However, the results would have been quite different if the Conventional Group, for example, had been the largest group—thus illustrating the influence of group size on outcome. In this respect, Option 2 provides a more stringent index of criterion-related validity. That is, the results for a large criterion group cannot overwhelm the results for the smaller groups. This problem can also be avoided by simply using criterion groups of equal size, as was done in an extensive study by Kuder (1968).

It is interesting to note in passing that one could achieve a hit rate of 67% under Option 1 simply by predicting membership in the largest group (the Social Group) for everyone. The weighted average hit rate for SDS raw scores was only 62%. This is just another instance of the base rate problem (Meehl & Rosen, 1955) resulting from application of the best *a priori* strategy

Table 1
Variation in Hit Rates Produced
by Two Validation Options

Occupational Preference Group	Sample Size ^a	Base Rate	Hit Rate	
			Raw Scores	Standard Scores
Investigative	139	14%	45%	52%
Artistic	117	12	49	46
Social	657	67	73	24
Enterprising	48	5	4	21
Conventional	24	2	29	54
Weighted average hit rate (Option 1)			62	31
Unweighted average hit rate (Option 2)			40	39

Note. Data for 432 women in a state liberal arts college and 557 women in a state university are summarized from a study reported by Gottfredson and Holland (1975).

^aData reported for the "Realistic" group were not analyzed because of an inadequate N of 4. Sample sizes for the Conventional and Enterprising groups are smaller than would be desired in a well-designed study.

(Cronbach & Gleser, 1965). Under Option 2, the hit rates for raw scores and standard scores "beat" the base rates.

Crucial Questions

Some may not agree that Option 2 is the appropriate procedure for interest inventory validation against group membership criteria. However, it should be clear that the purpose of assessment must be determined before deciding how to use group base rates in any study validating group membership predictions. It is not enough to say that the purpose of psychological science is to predict behavior. One must first determine which behavior it is appropriate to predict.

When one allows the relative sizes of criterion groups to influence the results of predictive validity studies (Option 1), one does not "let the chips fall where they may." A true indication of the weighted average hit rate of a measure cannot be obtained unless population base rates are

used instead of the base rates for the samples that happen to be at hand (Meehl & Rosen, 1955). Hence, criterion group base rates must be adjusted to reflect population base rates. Determination of the population base rates, however, can present serious problems. When validating an interest inventory via Option 1, should base rates be determined by number of workers per occupational category or by current employment needs? If the latter, should national or local needs be used? What about projected needs in five years? Clearly, a value judgment is involved. The decision to use Option 2 also involves a value judgment. In both instances, as in other areas of assessment, one must carefully examine the purpose of the measure being validated.

References

- Berdie, R. F. Vocational choice studies. *Personnel and Guidance Journal*, 1970, 48, 604-605.
Borgen, F. H. Predicting career choices of able college men from occupational and basic interest

- scales of the Strong Vocational Interest Blank. *Journal of Counseling Psychology*. 1972, 19, 202-211.
- Campbell, D. P. *Handbook for the Strong Vocational Interest Blank*. Stanford, CA: Stanford University Press, 1971.
- Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press, 1965.
- Dolliver, R. H. Concurrent prediction from the Strong Vocational Interest Blank. *Journal of Counseling Psychology*. 1975, 22, 199-203.
- 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, G. D., Holland, J. L., & Gottfredson, L. S. The relation of vocational aspirations and assessments to employment reality. *Journal of Vocational Behavior*, 1975, 7, 135-148.
- Holland, J. L. *Professional manual for the Self-Directed Search*. Palo Alto, CA: Consulting Psychologists Press, 1972.
- Kuder G. F. *General manual: Occupational Interest Survey*. Chicago, IL: Science Research Associates, 1968.
- Kuder, G. F. Some principles of interest measurement. *Educational and Psychological Measurement*, 1970, 33, 205-226.
- Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*. 1955, 52, 194-216.
- Prediger, D. J., & Cole, N. S. Sex-role socialization and employment realities: Implications for vocational interest measures. *Journal of Vocational Behavior*, 1975, 7, 239-251.
- Prediger, D. J., & Hanson, G. R. The distinction between sex restrictiveness and sex bias in interest inventories. *Measurement and Evaluation in Guidance*, 1974, 7, 96-104.
- Prediger, D. J., Roth, J. D., & Noeth, R. J. Career development of youth: A nationwide study. *Personnel and Guidance Journal*, 1974, 53, 97-104.
- Strong, E. K. *Vocational interests of men and women*. Stanford, CA: Stanford University Press, 1943.
- Zytowski, D. C. Predictive validity of the Kuder Occupational Interest Survey: A 12- to 19-year follow-up. *Journal of Counseling Psychology*, 1976, 23, 221-233.

Acknowledgment

The author is grateful to Robert Brennan, Senior Research Psychologist at ACT, for suggesting use of the terms "weighted (and unweighted) average hit rates" to distinguish between the two approaches to validation.

Author's address

Dale J. Prediger, Director, Developmental Research Department, Research and Development Division, The American College Testing Program, P. O. Box 168, Iowa City, IA 52240