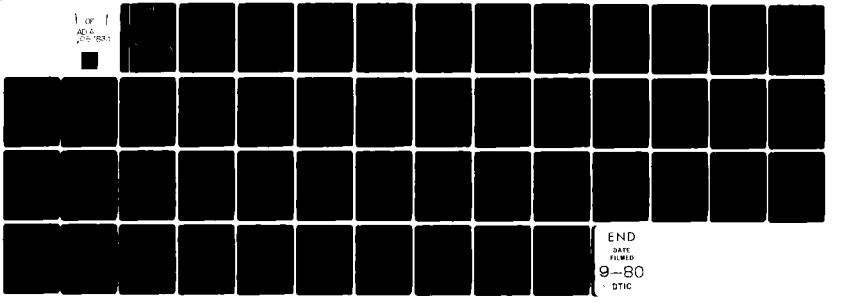
AD-A087 834 ILLINOIS UNIV AT URBANA-CHAMPAIGN DEPT OF PSYCHOLOGY F/G 5/10
APPLICATIONS OF ITEM RESPONSE THEORY TO ANALYSIS OF ATTITUDE SC--ETC(U)
JUL 80 C L HULIN, F DRASGOW, J KOMOCAR N00014-75-C-0904
UNCLASSIFIED TR-80-5 NL



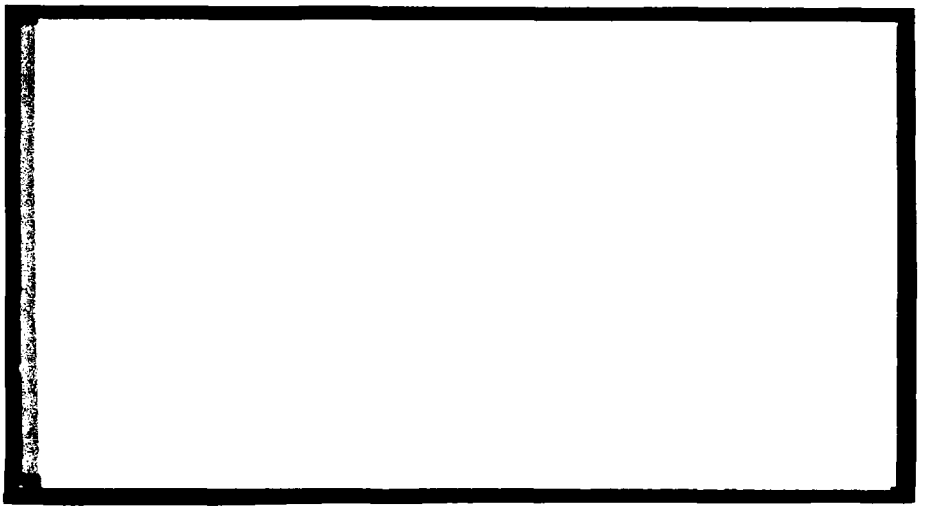
END
DATE FILMED
9-80
DTIC

LEVEL II

(6)

UNIVERSITY OF ILLINOIS

Studies of Individuals and Groups in Complex Organizations



AD A087834

DTIC ELECTE AUG 12 1980 D C

DDC FILE COPY

Department of Psychology
Urbana - Champaign

80 8 12 023

Applications of Item Response Theory To Analysis of Attitude Scale Translations

Charles L. Hulin
John Komocar
University of Illinois at Urbana-Champaign

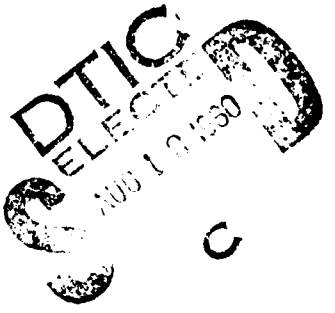
Fritz Drasgow
Yale University

Technical Report 80-5
July, 1980

Prepared with the support of the Organizational Effectiveness Research Programs, Office of Naval Research, Contract N000-14-75-C-0904, NR 170-802.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 80-5	2. GOVT ACCESSION NO.: AD-A087 834
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Applications of Item Response Theroy to Analysis of Attitude Scale Translations	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER:	
7. AUTHOR(s): Charles L. Hulin, Fritz Drasgow, John Komocar	8. CONTRACT OR GRANT NUMBER(s): N000-14-75-C-0904
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, University of Illinois, Champaign, IL 61820	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: N-170-802
11. CONTROLLING OFFICE NAME AND ADDRESS: Organizational Effectiveness Research Programs, Office of Naval Research (Code 452), Arlington, VA 22217	12. REPORT DATE: July, 1980
	13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

translations, item bias, Job Descriptive Index, Item Response Theroy, bilingual, cross cultural, latent trait, attitude measurement

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Methods of detecting item bias developed from a logistic item response theory (IRT) are generalized to analyze the fidelity of foreign language translations of psychological scales. These IRT methods are considered as alternatives to traditional sample dependent methods. Transformed item characteristic curves generated in the original and target languages, rather than item parameters from two languages, are examined for significance

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

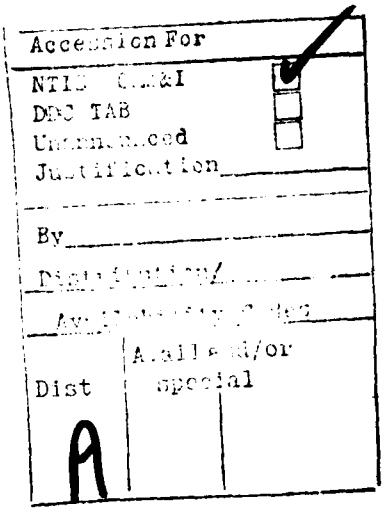
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

of differences. Data from a Spanish translation of the Job Descriptive Index are used to illustrate the method. It is argued that equivalent item characteristic curves across the original and translated items of a scale produce equivalent instruments in both languages, and nonequivalent item characteristic curves pinpoint differences between the two versions of the scale.

IRT Translations

Abstract

Methods of detecting item bias developed from a logistic item response theory (IRT) are generalized to analyze the fidelity of foreign language translations of psychological scales. These IRT methods are considered as alternatives to traditional sample dependent methods. Transformed item characteristic curves generated in the original and target languages, rather than item parameters from two languages, are examined for significance of differences. Data from a Spanish translation of the Job Descriptive Index are used to illustrate the method. It is argued that equivalent item characteristic curves across the original and translated items of a scale produce equivalent instruments in both languages, and nonequivalent item characteristic curves pinpoint differences between the two versions of the scale.



Information available to psychologists about individuals depends, ultimately, on comparisons for its meaning. Thurstone's Law of Comparative Judgment (1927) made one comparison process explicit: the method of pair comparisons. Here the observer judges which one of a pair of stimuli is greater with respect to a specified attribute for all possible pairs of stimuli in a stimulus set. From these data, the psychological values along the specified attribute (i.e., scale values) of each stimulus are determined and are based on the frequencies that stimuli were compared and confused with each other. A quite different comparison process is implicit in normative approaches to the study of individuals. Large numbers of randomly sampled individuals respond to the same or parallel instruments, thus providing group-based information about central tendency and variability. An individual's score becomes meaningful only when it is compared to the normative group's mean and standard deviation. In idiographic or idiothetic studies of individuals, the comparison process involves examining the responses of the focal individual across a large number of situations. Studies of person-situation interactions (Ekehammer, 1974) as well as three-mode factor analysis (Tucker, 1966) require the responses of many individuals to many stimuli across several situations. The attribute common to all these diverse research methods is that behavior becomes meaningful to the psychologist through comparison to other behaviors of the same or other persons in the same or other situations.

Individuals' scores on some measurement scales in wide use, such as IQ scales, may appear to have absolute meanings without reference to some explicit comparison process. Of course this is not true; frequent usage has

rendered the comparison process nearly automatic. Other scales that are used less frequently, such as the Miller Analogies Test, require a more apparent comparison process. Scores on a set of scales known and used only by a small segment of psychologists, such as the Job Descriptive Index (JDI) (Smith, Kendall, and Hulin, 1969), must be accompanied by information about means, variances, and normative populations before any meaning can be attributed to scores and differences between scores.

Two increasingly important areas of psychological research, particularly in applied psychology, are cross-cultural social psychology and its derivative, cross-national industrial-organizational psychology. Researchers in these areas study differences in interpersonal or organizational processes as they are influenced by the cultural and national settings of individuals or organizations. The need to assess and make statements about differences between cultures as well as within cultures is an integral part of this area of study. Herein lies the problem that we address in this paper: Cross-cultural research depends on cross-cultural comparisons, which, in turn, depend upon the meaningfulness of measuring instruments and scale scores both <u>within</u> and <u>between</u> the cultures in question. Scales must be meaningful within each culture as well as having the <u>same</u> meaning across both cultures in order for a comparison to yield useful information.

Our emphasis on cross-cultural measurement of psychological quantities does not mean that we are uninterested in cross-cultural comparisons based on non-equivalent scales. Such comparisons depend on two scales, each of which measures validly a given construct in one culture. For example,

relations of variables to measures reflecting satisfaction with interpersonal relations in a work situation may be made across cultures even though the two satisfaction scales may have quite different referents. In one culture, the scale may refer to satisfaction with the members of one's own work group, a very narrow referent, while in another culture the scale may refer to satisfaction with one's work group, supervisors, other members of the organization, and even individuals external to the organization (users of the product). Nonetheless, even with these different referents, statements about antecedents and consequences of satisfactions with interpersonal relations can be made within each culture that do not require knowledge about the scale used in the other. Individuals interpreting the differences in the relations obtained within the two cultures must be aware of the broader based measures used in one of the cultures. This does not invalidate the comparison.

Similarly, reliance by members of different cultures on qualitatively different types of information to make decisions, e.g., personnel decisions based on family names, religion, accent, or ability, allows measurement-free comparisons between cultures (Triandis, 1963). However, limiting cultural research to such scale-free data seriously handicaps researchers whose interests may include assessing relative amounts of affect across individuals, comparisons of changes in differences over time where the magnitudes of the differences as well as the trends are important, or even idiothetic studies of individuals drawn from two cultures.

In this way, cross-cultural research and attitude survey programs in multi-national firms share one very serious methodological problem. This

is, quite simply, the problem of translating an instrument developed in one culture and language into the language of the second culture, while preserving the integrity and meaning of the original instrument. Clearly, there are other problems along the path to psychometrically sound measurement of theoretically relevant variables. For example, deciding what variables to assess in the original and target language and whether to use centered or decentered translations (Werner and Campbell, 1970) are important in both the basic and applied research areas noted above.

One solution to the problem of choosing theoretically and culturally relevant variables involves two independent research programs, one in each culture. With a common research strategy, the two programs must proceed through the multiple steps of domain entry; eliciting reponses from individuals about the variables, events, actions, or actors that are important in determining their general attitudes toward important concepts; generating item universes defining each of the apparently relevant areas contributing to overall attitudes; analyzing items; determining/verifying dimensionality; developing psychometrically sound instruments to measure overall or specific attitudes; and collecting data to assess construct validity. Comparisons of the two sets of scales resulting from these two independent (except for the use of a common general strategy) procedures may reveal only modest communality between the two sets of instruments and the variables they measure. Direct comparisons of results based on one set of scales to results based on the companion set from the second culture can be made only at gross levels of semantic statements--there are positive relations between X and Y in both cultures-- and names given to scales and

factors-- satisfaction with pay, for example--must be viewed as shorthand labels summarizing consistencies the investigators see among the items composing each scale.

As an example, consider a set of satisfaction scales reflecting attitudes toward work and working conditions that might have been developed in Kibbutzim in Israel, in factories in Sweden, and in large organizations in the United States. Pay satisfaction very likely will have different meanings in the three cultures, referring to small group equality in Kibbutzim, broader based notions of equity involving national comparisons and government decisions and standards in Sweden, and individual perceptions of equity involving occupational and demographic comparison groups in the United States. Individuals' scores on each of the three scales might be related to withdrawal from their employing organizations. But scores on the three would likely have different roots requiring different intervention strategies if judged seriously low. Are we talking about the same construct in all three cultures? Perhaps. However, mathematical statements relating the three cannot be made; only imprecise semantic statements would be permitted. We may have achieved theoretical relevance within each culture at the expense of cross-cultural information.

An alternative approach begins with psychometrically sound instruments that have been developed in one culture, based on one language, and translates the existing set of scales into a new languaye that can be used in the target culture. Obviously, this procedure facilitates quantitative precision while raising questions of construct validity and construct relevance in the second culture. In this paper we propose methods for

studying the fidelity of translated scales.

By way of introduction to this approach, consider problems of psychological assessment in different ethnic samples in the United States. Our questions about cross-cultural comparisons and meaningfulness are closely paralleled by the theoretical and practical problems encountered when examining legal and statistical questions of test bias within the heterogeneous population of the United States. The analogy is all the more striking because the instruments used in different samples within the United States were probably developed on homogeneous samples of English speaking members from our population. In traditional approaches, the determination of bias usually proceeds in a step by step, hierarchical fashion. The examination of potential test bias usually follows a finding of mean differences between samples drawn from different ethnic, race, or sex groups. The relation between the test or scale and some external criterion, presumed to be related to the construct being measured, is then determined. Comparison of the demands of one of the definitions of test bias to the empirical relations between the scale and the criterion within the two subpopulations usually yields evidence about the fairness of the test or scale.

Such a hierarchical procedure is not without difficulties. Selection of an unbiased criterion against which to judge the scale, relevance of the criterion to the scale being examined, and even choice of internally consistent definitions of bias or unfairness (Peterson and Novick, 1976) are matters of controversy. Finally, given the usually weak relation between the scale in question and an external criterion, apparent lack of bias may

be found because equally weak relations between experimental scale and criterion are generated by much different psychological processes within the two different ethnic or racial samples.

Examination of test bias from the perspective of item response theory (IRT) would proceed quite differently. In an IRT-based approach to test development, the item characteristic curve (ICC) is fundamental. The ICC displays the conditional probabilities of passing an item (getting the item correct or giving a positive response) for each value of the assumed underlying trait or ability (θ). ICC's can be used for test development as well as examining biases for or against members of identifiable sub-groups of the population. Thus, in IRT, bias can be defined in terms of the ICC's for different subsamples: (equated) θ's in the different groups yield different conditional probabilities of passing the item or making a positive response.

It is important to note a fundamental difference between traditional approaches and IRT in assessing bias. In traditional approaches, bias is examined by comparing the relations between the test or scale and an external criterion across two subpopulations. For example, according to one definition of test bias, a test is biased if there are unequal slopes in the regressions of criterion onto test score for two subpopulations (Cleary, 1968). Thus, a test is biased or not only with respect to some external criterion. In IRT, bias is not generally assessed with respect to an external criterion. Instead, item bias is assessed by examining the relation between the conditional probabilities of passing an item (given θ) and θ, the unidimensional latent trait measured by the item. Thus, ICC's

for the two subpopulations are compared. Note that in IRT, no criterion data are required; ICC's are estimated from the responses to items composing the scale of interest. Bias, or its lack, is judged relative to the underlying trait, not to an arbitrary external criterion.

We are not asserting that traditional approaches to item or test bias have relied exclusively on differential relations between scales and external criteria. Some research has been conducted using internal criteria. These internal criteria have usually been differences in item total biserial correlations computed within different groups, or item difficulty by group interactions (Angoff & Ford, 1973; Green & Draper, Note 1; Ironson & Subkoviak, 1979). The greatest emphasis in recent studies has been on relations with external criteria.

Reliance on a purely internal criterion, an ICC, to detect bias is a strong position and cannot be defended in the extreme situaion where all of the items composing a test are biased in the same direction and by the same amount in one of the sub-samples. Although such an occurence is theoretically feasible, the probability seems small. Such an occurrence would, however, generate a test that was biased but gave the appearance in IRT of being unbiased. Similarly, the items of a test or scale could be unbiased with respect to θ in both samples but the collection of items could be assessing an underlying trait that was necessary for performance in one culture or sample but was irrelevant in another culture or sample. This state of affairs would be detected as biased by traditional criterion referenced approaches but not by IRT. It is important to note, however, that IRT gives an investigator two chances to detect test or scale bias:

once by examining item bias by ICC's and once by using external references to examine scale/criterion relations for bias.

If IRT were applied to translation procedures, the ICC's generated by the different translations of the same item could be used to provide evidence about the quality of the translation from the original to the target language, about the meanings of the items relative to the underlying trait being measured, and about the equivalence of test or scale scores across the two languages and cultures.

Translation of psychological measuring instruments into new languages involves a series of steps. First, translation into the target language and back translation into the original language by multiple independent translators is required. This is simply a check and verification on the general quality of the translation and should be done for any translation. Its importance cannot be overstated. Lack of convergence back into the original language is apparent, and remedial action can be achieved at this point by refining problem items. This procedure is necessary but not sufficient for generating equivalent scales. Convergence, it should be noted, can be achieved by highly skilled translators who translate from the target language back into the original: garbled translations can be translated into a close approximation of the original by insightful guesses and inference. Thus, fidelity of the translation to the original is not guaranteed by convergence.

The second and following steps differentiate more sharply between item response theory and other procedures. As a step prior to any construct validation, a frquently used method would normally obtain a sample of

bilingual subjects similar to those who would eventually complete the test or scale. These subjects complete the test twice, once in the original language and once in the target language. Statistics summarizing the data from the two versions provide the basic comparisons of interest. Means, variances, and item-item covariances are compared, differences noted, and frequently items are reworded and another iteration is attempted. Following this, prediction of an external criterion or other empirical validation procedure would be used (Irvine & Carroll, 1980).

A more sophisticated version of this procedure was provided by Katerberg, Hoy, and Smith (1977) in their analysis of a translation of the JDI (Smith et al. 1969) into Spanish. Bilingual employees of a large organization were administered both versions of the JDI twice in a counter-balanced order separated by 30 days. Katerberg et al. analyzed these data using the outlines of Cronbach's generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). The authors estimated variance due to language, time, subjects, and interactions. In addition, equations transforming scores from Spanish to English were provided to permit the use of English language norms with Spanish language data. Their substantive conclusions were generally positive and they concluded they had a good translation of the JDI. Proportions of variance due to persons ranged from .59 to .68 across the five scales; person by time interactions, which could be either true change or error, accounted for between 22 and 31 percent of the scale variances; error variance indicated in person by language, and person by language by time interactions accounted for between 00 and 13 percent of the scale variance.

The limitations on such analyses are obvious. Samples of bilingual subjects similar to those who will eventually respond to the scales are required. Such samples are not always available and, even when available, bilingual subjects may differ substantially from monolingual subjects in terms of their semantic structures and the subtle shadings of differences they see among words in their languages. Thus, results may or may not be generalizable to other samples from populations of interest.

Because the analysis of translated scales using a traditional approach is normally done on scale scores rather than item responses, it is entirely possible that analysis of translation data will reveal scale displacements (unequal means) or unequal units of measurement (variances) that will require adjustment of sample estimates before any direct comparisons across languages can be made. Equations must then be provided to transform responses to the translated version of the scale into the metric of the original language. Finally, serious discrepancies in means or variances will be revealed only in the units of the analysis, in this case in terms of scale scores, and no indications are provided concerning which items must be retranslated or adjusted in order to achieve equivalence.

The IRT approach to analysis of translation data that we propose proceeds much differently. Here, ICC's generated by the same item in two different languages provide direct evidence about the meanings of the items in terms of the underlying latent trait being measured by each version of the scale. ICC's for an item that differ across languages (after equating metrics) pinpoint those items in need of revision, suggest the type of revision (e.g., more or less difficult), indicate items with different

discriminating power, and may even reveal problems with lower asymptotes (in terms of multiple choice tests) resulting from ineffective or overly seductive distractors. Further, because the analysis of the scales is in terms of observable data at the item level, similar ICC's across all items automatically results in tests with similar norms in both languages. Scale scores can then be interpreted using available norms from the original language and the necessity for providing equations to equate scores from the different versions is removed.[1]

In the present study, we reanalyze the JDI data reported by Katerberg et al. (1977) to illustrate the strengths and weaknesses of one IRT approach for conceptualizing and analyzing the quality of a translation. We believe this procedure allows a more penetrating examination of the items composing the JDI than does generalizability or traditional approaches. Unfortunately, these benefits result, in part, from making stronger assumptions, which appear to be violated. Throughout the remainder of this paper we indicate the advantages of the IRT approach and the assumptions that our data may violate.

Analysis

Subjects, Measures, and Data

The following paragraphs present a brief description of the data analyzed herein; a fuller description is presented by Katerberg et al. The original sample consisted of 203 bilingual employees of a large merchandising firm. They were of Cuban or Puerto Rican extraction and employed in company units in the Miami or New York City areas. Sales, sales support, and supervisory functions were represented in the sample.

Respondents were asked to complete both English and Spanish versions of an attitude questionnaire on two different occasions 30 days apart in a counterbalanced order. Questionnaires with greater than 10% missing data on either the English or Spanish versions of the scales of interest to the present study were deleted. This criterion reduced the sample for the present study to 178 useable questionnaires. A total of 173 useable questionnaires were taken from the first data collection period and five were taken from the second data collection period. The latter five questionnaires were from subjects with excess omitting during the first data collection session but with useable data from the second session.

The scales examined herein include the English and Spanish versions of the 72 item, five scale (measuring satisfactions with the work itself, pay, promotional opportunities, supervisor, and coworkers) JDI developed by Smith et al. (1969) to assess attitudes of workers in a wide variety of organizations. In particular, the original English versions of these scales are compared to their Spanish translations to evaluate the quality of the translation. Note that a high quality translation would allow comparisons of mean levels of job satisfaction for English speaking and Spanish speaking workers. In addition, broader questions of the meanings of work within different cultural groups could begin to be studied.

Analysis

The two parameter logistic model (Birnbaum, 1968) was selected as a statistical model for the JDI items. In this model, the probability of an affirmative response to the ith JDI item, given a satisfaction level of θ, is

$$\text{Prob (Positive Response } | \theta) = \frac{1}{1 + \exp [-D\, a_i(\theta - b_i)]}$$

Here b_i corresponds to the point on the θ continuum where the probability of an affirmative response from a randomly selected worker is .5, a_i controls the steepness of the ICC, and D is a scaling factor for logistic approximation to the normal ogive model usually set equal to 1.702. In the context of mental tests, a_i and b_i correspond, respectively, to item discriminating power and item difficulty. In attitude assessments, they refer to discriminating power and extremity of item wording, respectively.

Our primary interest in the present research is in comparing the equality of ICC's for the English version of JDI items to the corresponding ICC's for the Spanish version. At present, however, distribution theory for estimated ICC's has not been derived. Thus, a straightforward test of the equality of ICC's is not possible. We have developed the following heuristic procedure for comparing ICC'S.

The initial step involves separate maximum likelihood estimation of item and person parameters for the English and Spanish data sets. The LOGIST computer program, developed by Lord and his colleagues (Wood, Wingersky & Lord, Note 2; Wood & Lord, Note 3) can be used for this purpose. Since the parameters (and parameter estimates) of the two parameter logistic model are not uniquely determined, it is necessary to equate metrics. A procedure, such as the one developed by Linn, Levine, Hastings and Wardrop (Note 4), would normally be used to equate metrics. However, a more direct procedure is possible for the present data because the bilingual subjects

completed questionnaires in both English and Spanish. Thus, for each subject we can plot the estimate of job satisfaction from the Spanish version ($\hat{\theta}_S$) against the estimate of job satisfaction from the English version ($\hat{\theta}_E$). This plot is shown in Figure 1. The correlation between $\hat{\theta}_S$ and $\hat{\theta}_E$ is .92 and the regression of $\hat{\theta}_S$ on $\hat{\theta}_E$ is $\hat{\theta}'_S = -.01 + .96\ \hat{\theta}_E$. In light of the Monte Carlo research studying the standard error of estimate for θ (Lord, Note 5; Swaneriathon & Gifford, Note 6), it appears that no adjustment of the theta metrics is necessary. Figure 2 further confirms this conclusion. Here estimated "item difficulties" for the English and Spanish versions of the JDI are plotted. The correlation between the two sets of estimated item difficulties is .93. and the regression of estimated item difficulty for Spanish items ($\hat{b}_S$) on estimated item difficulty of English item ($\hat{b}_E$) is $\hat{b}'_S = .01 + 1.02\ \hat{b}_E$.

Insert Figures 1 and 2 about here

Having determined that the θ_S and θ_E metrics are equivalent, we can now begin to compare ICC's. An indirect test of the equivalence of ICC's can be performed by obtaining "empirical ICC'S." An empirical ICC is computed by first dividing the $\hat{\theta}$ continuum into a number of mutually exclusive and exhaustive intervals. Then the proportions of positive responses from subjects within the intervals are determined. The plot of these proportions against the corresponding center of each $\hat{\theta}$ interval constitutes an empirical ICC.

Figure 3 presents empirical ICC's for the Spanish and English versions

of the item "challenging." These curves were obtained by dividing the θ continuum into 18 intervals with approximately 10 respondents per interval. We selected 18 intervals and 10 subjects per interval because this seemed to be the most reasonable trade-off between (1) a large number of points on the empirical ICC; and (2) accurately estimating each point of the empirical ICC. Obviously, more points, more accurately estimated, would be desirable.

Insert Figure 3 about here

To compare empirical ICC's, the proportions of positive responses were transformed by a logit transformation. The rationale for this transformation can be seen by examining its effects on the theoretic ICC. Here

$$L\{P_i(\theta)\} = \log\left[\frac{P_i(\theta)}{1 - P_i(\theta)}\right]$$

$$= \log\left[\frac{\{1 + \exp[-Da_i(\theta - b_i)]\}^{-1}}{1 - \{1 + \exp[-Da_i(\theta - b_i)]\}^{-1}}\right]$$

$$= Da_i(\theta - b_i)$$

and thus $L\{P_i(\theta)\}$ is a linear function of θ. The empirical proportions, $\tilde{P}_i(\hat{\theta}$ Interval) also become linearly related to $\hat{\theta}$ after the logit transformation is applied. Finally, the regressions of $L\{\tilde{P}_i$ ($\hat{\theta}$ Interval)$\}$ onto $\hat{\theta}$ can be computed for the English and Spanish versions of the JDI items and a

statistical test of their equivalence can be carried out (Neter & Wasserman, 1974, pp. 161-167). A significant difference may be interpreted as indicating nonequivalence of ICC's across English and Spanish versions of the item. In addition, a significant effect can be examined more closely to determine whether the slopes of the regression lines differ (which would imply a difference in the two a parameters) and whether the intercepts differ (if the slopes do not differ, then significantly different intercepts imply a difference in the b parameters).

Table 1 presents a summary of the significance test for the 72 item JDI. Of the 72 F-ratios calculated (df = 2,16), three were significant at α = .05. Taken alone, these results indicate that three Spanish JDI items have ICC's that differ from the ICC's of the corresponding English JDI items. However, we note that these results could be Type I errors. Table 1 presents the obtained and expected (under H_0) numbers of significant F-ratios for selected α-levels from .01 to .50. There is little difference between the obtained and expected numbers of significant F-ratios at α-levels of .01, .05, and .10. Taken in total, the data in Table 1 indicate a very good translation of the JDI. The three Spanish-English ICC pairs that were significantly different at an α-level of .05 should be independently verified in a new data set before we conclude bias is present.

Insert Table 1 about here

Figures 4 and 5 provide graphic examples of our regression method for comparing the equality of ICC's. In Figure 4, the transformed, empirical

ICC's of the English and Spanish versions of the item "challenging" are presented. It should be emphasized that an error in transcribing the translation of the JDI items resulted in the item "challenging" being rendered as "retador" rather than its equivalent "desafiante." This error in translating was appropriately detected by these ICC analyses as generating a biased item. Traditional analysis of total scores would not be able to detect errors at the item level.

Insert Figure 4 about here

Figure 5 shows a similar comparison for the item "influential." Conclusions based on visual inspections of the regression lines in each of the figures agree with significance tests for the equality of the regression lines. The regression line for the item "challenging" in English differs from the regression line for the corresponding item in Spanish ($F = 16.5$, $p < .05$) and the regression lines for the item "influential" do not differ significantly ($F = .13$, $p > .05$). Visual inspection of the dispersion of the logit transformed conditional probabilities, $L(\tilde{P}_i$ ($\hat{\theta}$ Interval)$)$, about the regression lines also reveals, as previously mentioned, that more points, more accurately estimated would be desirable.

Insert Figures 4 and 5 about here

Discussion

One of the critical assumptions of IRT is that the latent trait space is unidimensional. This assumption is difficult, very likely impossible, to make with any degree of assurance based on an examination of a real data set. Except in the degenerate case of a one item scale, the assumption is probably never strictly true with real data. The assumption seems even more tenuous in the case of the JDI because factor analyses of the scales by the original developers (Smith, Kendall, and Hulin, 1969) and others (Smith, Smith, and Rollo, 1974) have repeatedly concluded that the 72 items indeed generate five separate dimensions that are defined by the intended items. In fact, the evidence for multidimensionality is probably as well established for this set of items as it is for any comparable psychological instrument.

Very recently, however, investigators have presented data from more complex analytic techniques suggesting that treating structures of the JDI as comprising a large general affect component, an "A" dimension analogous to "G" in ability, and five group factors corresponding to the five scales may not only be warranted by the data (Parsons and Hulin, Note 7) but may even improve the psychometric properities of the instrument for estimating an underlying construct (Drasgow and Miller, in press). Specifically, Parsons and Hulin first treated the 72 items as if they were unidimensional and estimated and the IRT a_i and b_i parameters for each of the items using LOGIST. The resulting a_i parameters of the items were related to the loadings of the 72 items on a general "A" measure derived from a hierarchical factor analysis.[2] The a_i parameters of the 72 items correlated

approximately .80 with the loadings of the items on the general factor. A similar relation was obtained between the a_i parameters and the loadings of the items on the first unrotated principal component.

These results based on responses to the JDI seem analogous to Lord's work on the SAT-V scale (Lord, 1968). At one level of analysis, the SAT-V items factor rather cleanly into one dimension with all items loading on that one dimension. Nevertheless, it appears possible to refine this dimensionality further by factoring the items within the space defined by the verbal items and extract four or five meaningful dimensions reflecting performance on clusters of items assessing verbal analogies, reading comprehension, antonyms, etc.

It is possible, although for those who prefer a world with a minimum of ambiguity it may not be satisfactory, that answers to questions about unidimensionality or multidimensionality depend on our purposes. Many early studies using the JDI as a multivariate instrument have provided evidence that the five scales are not redundant when treated as five separate dimensions in empirical research: each provides some important, scientifically meaningful unique variance (Adams, Laker, and Hulin, 1977; Herman, Dunham, and Hulin, 1975). It is nevertheless true that the general "A" factor seems to account for the bulk of explainable variance in absenteeism, turnover, and other variables (Miller, Note 8; Drasgow & Miller, in press).

It is interesting to pursue further the analogy between the JDI and the SAT-V and examine the consequences of treating the apparently multidimensional JDI as unidimensional or treating the SAT-V, long

considered unidimensional, as spanning a 4 or 5 dimensional latent trait space. At this time we know of no data suggesting that treating the SAT-V as multidimensional leads to better predictions or understanding of undergraduate scholastic performance. It appears that the utility of assessing factors or dimensions beyond a powerful general first factor or dimension can best be evaluated within the context of a particular application. In a study that did treat different SAT-V item types as separate dependent variables, Alderman and Powers (Note 9) found that significant gains in SAT-V scores were primarily due to coaching effects on analogy and antonym items. The parallel between the Alderman and Powers coaching study and the Adams, Laker, and Hulin (1977) study using the JDI as a set of criterion scales is striking.

Translating the JDI also seems to be an example where the unidimensionality assumption is useful and does not do gross violence to the data. Here IRT offers an attractive procedure for examining item bias within psychological scales and tests. In the research reported here, the JDI seems to have been translated adequately--at least within the limits imposed by a priori α-levels and 72 simultaneous comparisons on correlated variables. We emphasize that these data and the analyses were intended as a demonstration of the applicability of the technique. Weaknesses of the present data set include the small sample size and bilingual subjects whose semantic structures differ significantly from mono-lingual Hispanics who might be expected to respond to these translated scales. Note, however, that these data have been analyzed previously from the perspective of generalizability theory, which makes the IRT results more interesting

because they could be compared to the results generated by Katerberg, Hoy, and Smith (1977).

These previous investigators concluded that the amounts of variance introduced into the scales by the translation process were small because the estimates of variance introduced by language differences were zero across all five scales of the JDI. Correlations between different language versions of the JDI scales ranged from .82 to .92. Nevertheless, when they tested the hypothesis that the regression equations of JDI English scores, for example, onto Spanish JDI scores had intercepts of zero and slopes of 1.00, the hypothesis had to be rejected. Thus, even with very small standard errors of estimate, Katerberg, et al. concluded that an equation would have to be developed to transform scores from one language into the metric of the other language to allow the use of the English norms for interpretation purposes.

Our conclusions, subject to the restrictions already discussed, would be that the 72 items do not appear to contain bias except in the case of the translation of "challenging," the theta metrics of the translated scales appear equivalent, and given these two conditions, translations via equations from one set of scores to another in order to allow the use of English norms is not necessary.

Within these limitations, the empirical results and conceptual developments are promising. Generalization must be cautious, however. The next step in verifying our empirical results is to administer the JDI to a large monolingual sample of Spanish speaking people. ICC's based on the very large number of English speaking workers who have responded to the JDI

can be compared to the ICC's generated by the monolingual respondents. If this procedure generates a number of ICC's that are significantly different and if these items are the same ones identified in this investigation as having deviating ICC's, the revised conclusions will be that three or four of the items need to be retranslated. If those items generating the expected number of significantly different ICC's from the monolingual sample are a different set from those identified in these data, we would conclude we are observing α-levels in action and adopt the translation.

Some further limitations on our analysis are apparent. We can conclude only that the translated items are unbiased and estimating equivalent traits within the two languages. The procedure is strictly an internal analysis and statements about equivalence must be made with this in mind. We have not examined relations between θ's derived from the translated scales and external variables nor have we studied the location of the translated scales within networks of relation derived in Spanish language cultures. It must be noted that how the latent trait functions in the second culture remains to be determined, but radically different meanings of the scales across the two cultures and languages would not be a fault of the translation. Instead, it might be attributed to a lack of cultural relevance. For example, the items composing the scale assessing satisfaction with the work itself on the job seem to have been successfully translated; the items generate equivalent ICC's in both languages. If, however, in a Spanish culture, satisfaction with the work itself was of little consequence to workers, certainly not something worth quitting or being absent about, then the two scales might have different behavioral correlates in the different

cultures.

Note that to argue that two scales have different meanings in two cultures on the basis of different behavioral correlates of the scales in the two cultures implies that we adopt an epistemological position that attributes meaning to variables in terms of their relations with other variables, whose meanings also must be inferred from their relations with still other variables, etc. ad infinitum. Definitional and semantic legerdemain will not provide solutions to our problems except as the definitions are terms imbedded within theories suggesting useful variables to include in the defining networks.

To conclude that the original and translated scales in two different languages are equivalent on the basis of similarity of ICC's implies that this form of internal consistency (not necessarily that assessed by coefficient α's, KR-20, or factor analysis) is sufficient to allow one to claim equivalence. This is a strong conclusion. Whatever epistemological position adopted, applications of IRT to translation problems in psychology does eliminate the necessity for bilingual samples with their different semantic structures and different interpretations of constructs-- different from either group of monolingual individuals with whom they share one language.

Our purpose in this article was not to introduce IRT applications to the analysis of translated scales as an alternative to examining the construct validity or empirical meaning of such translated scales in the relevant cultures. Our purpose was more modest in scope. We have presented the IRT analysis as an alternative step in the generation and analysis of

high quality translated scales. It obviously cannot be substituted for careful translation and back translations. Nevertheless, such IRT analyses obviate the necessity for obtaining samples of bilingual subjects who must respond to the translated scales as well as the original scales.

Methodologically and conceptually, the applications of IRT seem to stand outside a continuum ranging from evaluating an instrument against a single criterion (as might be done in a test fariness study) and the laborious and time consuming construct validations procedures outlined by Irvine and Carroll (1980). It provides more and better evidence about item bias than do test fairness procedures, but provides less information about the empirical meaning of the scales in the two cultures than a construct validation procedure. Although we disagree with many of the particulars outlined by Irvine and Carroll, we are in agreement with the necessity and purposes of construct validation of the scales in the two cultures.

Establishing equivalence of scales in two languages and cultures is clearly difficult. At the extreme it may involve simultaneous development of extensive, fully articulated nomological networks in both cultures. This procedure, by itself, is difficult. However, the crux of the problem is that the meanings of the variables most central to each network do not emerge unbidden from the background of quantitative and theoretical relations. The meanings depend as much on observers' fallible judgements, common sense, intuition, filtration of information, and hundreds of unproved assumptions as they do on objective descriptions of the relations obtained (Roberts, Hulin, & Rousseau, 1979). Different social scientists bring different intellectual backgrounds to bear on results of construct

validation efforts. Different interpretations of the meanings of the variables will result. We have attempted to provide a method that will circumvent some of the problems that must be solved before such nomological networks establishing construct validity can be developed.

Footnotes

This research was supported by ONR Contract N000-14-75-C-0904 Charles L. Hulin, Principal Investigator. The authors would like to thank Frank J. Smith for providing the sample of bilingual subjects and Robert Linn, Neil Dorans, Malcolm Ree, and Harry Triandis for reading and commenting on previous drafts. We would also like to thank Michael V. Levine for suggesting that we apply the logit transformation to the empirical ICC's.

Requests for reprints should be sent to Charles L. Hulin, Department of Psychology, University of Illinois, Champaign, Illinois, 61820.

1. This, of course, assumes near perfect statistical power and ability to detect small differences. Less than perfect detectability could result in small but consistent differences at the item level that accumulate to produce suspect scores.

2. This particular hierarchical analysis extracted five factors, rotated the factors obliquely, and extracted a second order general affect dimension that accounted for the obliqueness of the five first order factors. This second order factoring was followed by a Schmid-Leiman (1957) transformation that reorthogonalized the original five factors and expressed the general and the five group factors in terms of item loadings.

References

Adams, E.F., Laker, D.R., & Hulin, C.L. An investigation of the influences of job level and functional specialty on job attitudes and perceptions. Journal of Applied Psychology, 1977, 62, 335-343.

Angoff, W.H. & Ford, S.F. Item-race interactions on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10, 95-105.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. & Novick, M.R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Cleary, T.A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Drasgow, F., & Miller, H.E. Psychometric and substantive considerations in scale construction and validation. Journal of Applied Psychology, in press.

Ekehammer, B. Interaction in personality from an historical perspective. Psychological Bulletin, 1974, 81, 1020-1048.

Herman, J.B., Dunham, R.B., & Hulin, C.L. Organizational structure, demographic characteristics, and employee responses. Organizational Behavior and Human Performance, 1975, 13, 206-233.

Ironson, G.H. & Subkoviak, M.J. A comparison of several methods of assessing item bias. Journal of Educational Measurement, 1979, 16, 209-225.

Irvine, S.H. & Carroll, W.K. Testing and assessment across cultures: Issues in methodology and theory. In H.C. Triandis and J.W. Berry (Eds.). Handbook of cross-cultural psychology (Vol. 2). Boston: Allyn and Bacon, 1980.

Katerberg, R., Hoy, S., & Smith, F.J. Language, time, and person effects on attitude scale translation. Journal of Applied Psychology, 1977, 62, 385-391.

Lord, F.M. An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lord, F.M. & Novick, M.R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Neter, J. & Wasserman, W. Applied linear statistical models. Regression, analysis of variance, and experimental design. Homewood, Illinois: Irwin, 1974.

Peterson, N.W. and Novick, M.R. An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 1976, 13, 3-29.

Roberts, K.H., Hulin, C.L., & Rousseau, D.M. Developing an interdisciplinary science of organizations, San Francisco: Jossey-Bass, 1979.

Schmid, J. & Leiman, J. The development of hierarchical factor solutions. Psychometrika, 1957, 22, 53-61.

Smith, P.C., Kendall, L.M., & Hulin, C.L. Measurement of satisfaction in work and retirement, Chicago: Rand McNally, 1969.

Smith, P.C., Smith, O.W., & Rollo, J. Factor structure for blacks and whites of the Job Descriptive Index and its discrimination. Journal of Applied Psychology, 1974, 59, 99-100.

Thurstone, L.L. A law of comparative judgement. Psychological Review, 1927, 34, 273-286.

Triandis, H.L. Factors affecting employee selection in two cultures. Journal of Applied Psychology, 1963, 47, 89-96.

Tucker, L.R., Some mathematical notes on three mode factor analysis. Psychometrika, 1966, 21, 279-311.

Werner, O., and Campoell, D.T. Translating, working through interpreters and the problem of decentering. In R. Carroll and N. Cohen, (Eds.). A handbook of method in cultural anthropology. New York: American Museum of Natural History, 1970, 398-420.

Reference Notes

Note 1

Green, D.R. & Draper, J.F. Exploratory studies of bias in achievement tests. Paper presented at the Annual Meeting of the American Psychological Association, Honolulu, September, 1972.

Note 2

Wood, R.L., Wingersky, M.S. & Lord, F.M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, N.J.: Educational Testing Service, June, 1976.

Note 3

Wood, R.L. & Lord, F.M. A user's guide to LOGIST. Research Memorandum 76-4. Princeton, N.J.: Educational Testing Service, May, 1976.

Note 4

Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. An investigation of item bias in a test of reading comprehension (Technical Report No. 163). Illinois: University of Illinois at Urbana-Champaign, Center for the Study of Reading, March, 1980.

Note 5

Lord, F.M. Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. Research Memorandum 75-33. Princeton, N.J.: Educational Testing Service, 1975.

Note 6

Swaneriathon, H. & Gifford, J.A. Estimation of parameters in the three-parameter latent trait model. Laboratory of Psychometric and Evaluation Research Report No. 90. Amherst, MA: School of Education, University of Massachusetts, 1979.

Note 7

Parsons, C.K. & Hulin, C.L. An empirical comparison of latent trait theory and hierarchical factor analysis in applications to the measurement of job satisfaction (Technical Report 80-2). Illinois: University of Illinois, Department of Psychology, March, 1980.

Note 8

Miller, H.E. Withdrawal behavior among organization members. Unpublished doctoral dissertation. University of Illinois, 1980.

Note 9

Alderman, D.L. & Powers, D.E. The effects of special preparation on SAT-Verbal scores. Research Memorandum 79-1. Princeton, N.J.: Educational Testing Service, 1979.

TABLE 1

Summary of Significance Tests Comparing

ICC's of Corresponding English and Spanish Items

Nominal Alpha Level	Number Significant	Expected Number Significant
.01	1	.7
.05	3	3.6
.10	7	7.2
.25	12	18.0
.50	24	36.0

Note: There are a total of 72 JDI items.

Figure Captions

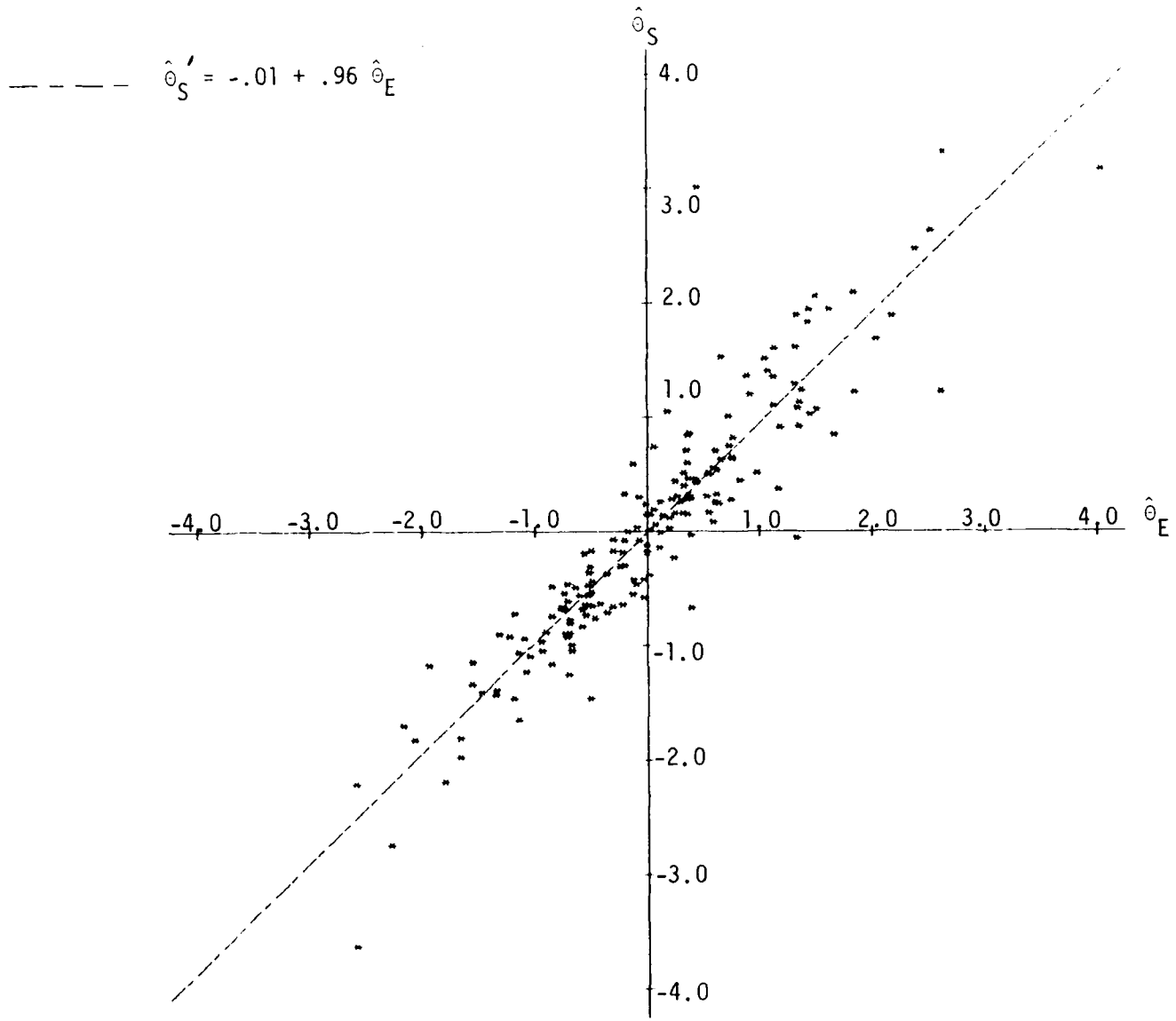
Figure 1. Regression line and scatter plot for regression of LOGIST estimated θ's from Spanish version of JDI on LOGIST estimated θ's from English version of JDI.

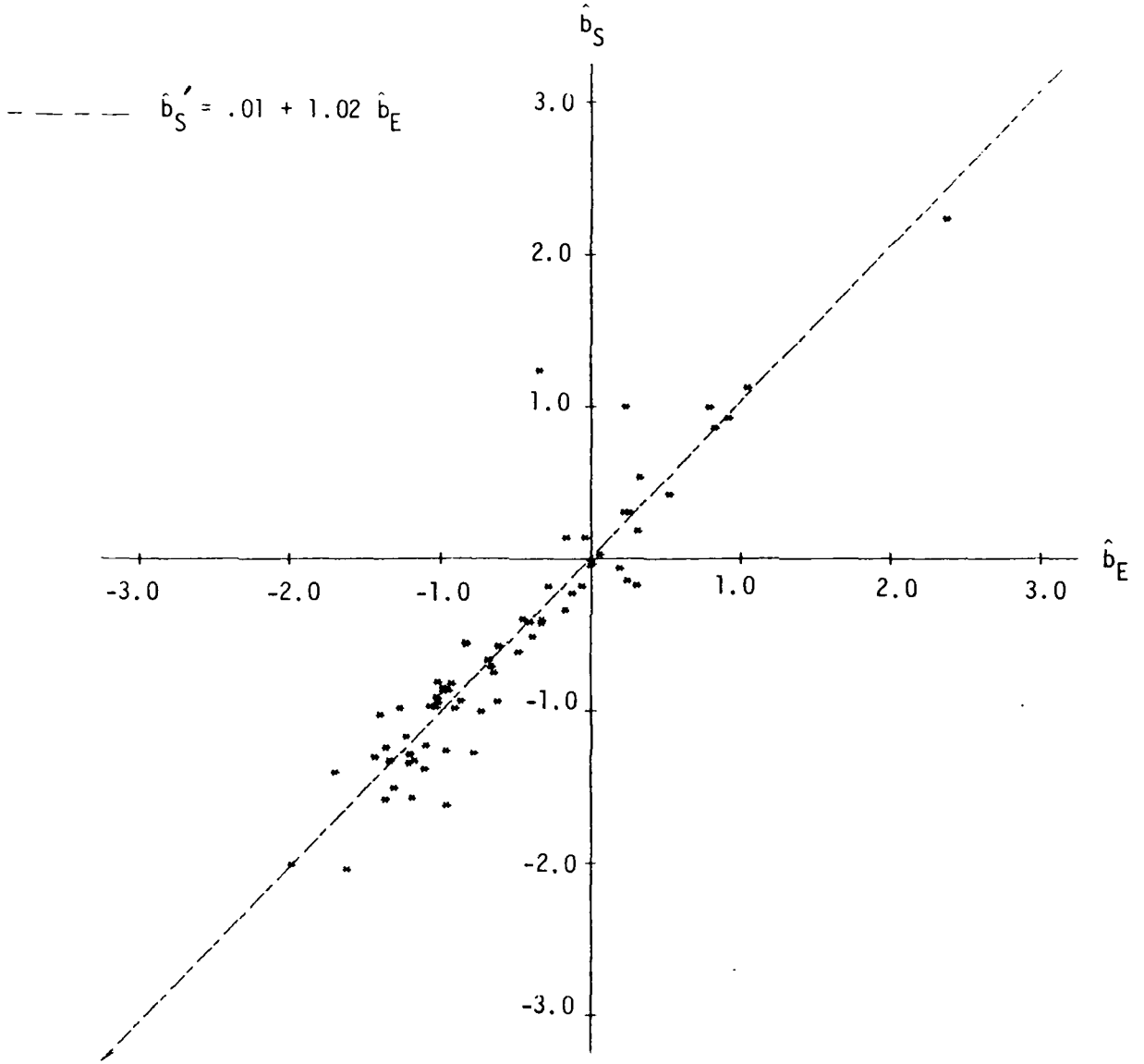
Figure 2. Regression line and scatter plot for regression of LOGIST estimated b item parameters for Spanish version of JDI on LOGIST estimated b item parameters for English version of JDI.

Figure 3. Proportion of Affirmative Responses for 18 $\hat{\theta}$ intervals to English and Spanish Versions of the JDI item "Challenging" and corresponding LOGIST estimated ICC's.

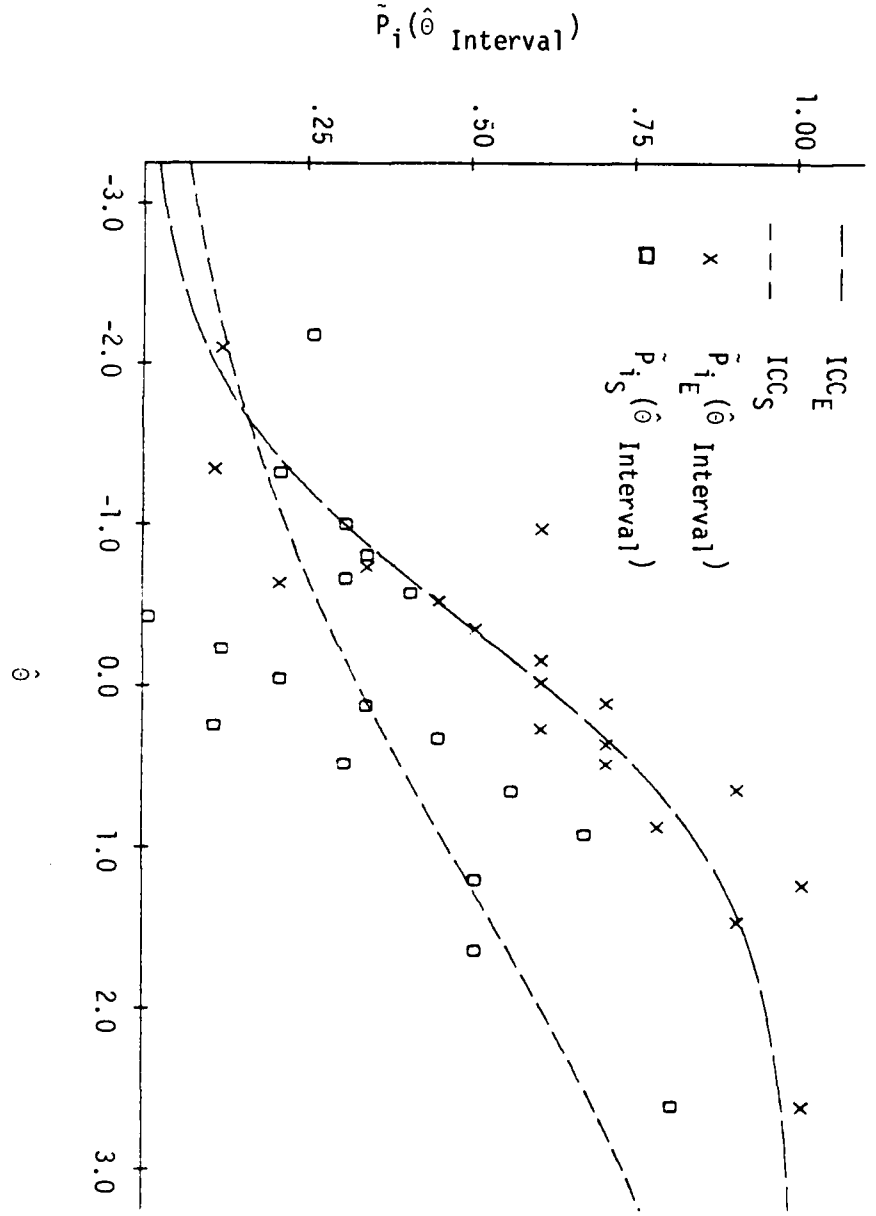
Figure 4. Regression lines and scatter plots for regression of logit transformed proportions of affirmative responses for 18 $\hat{\theta}$ intervals on $\hat{\theta}$ for English and Spanish versions of the JDI item "Challenging."

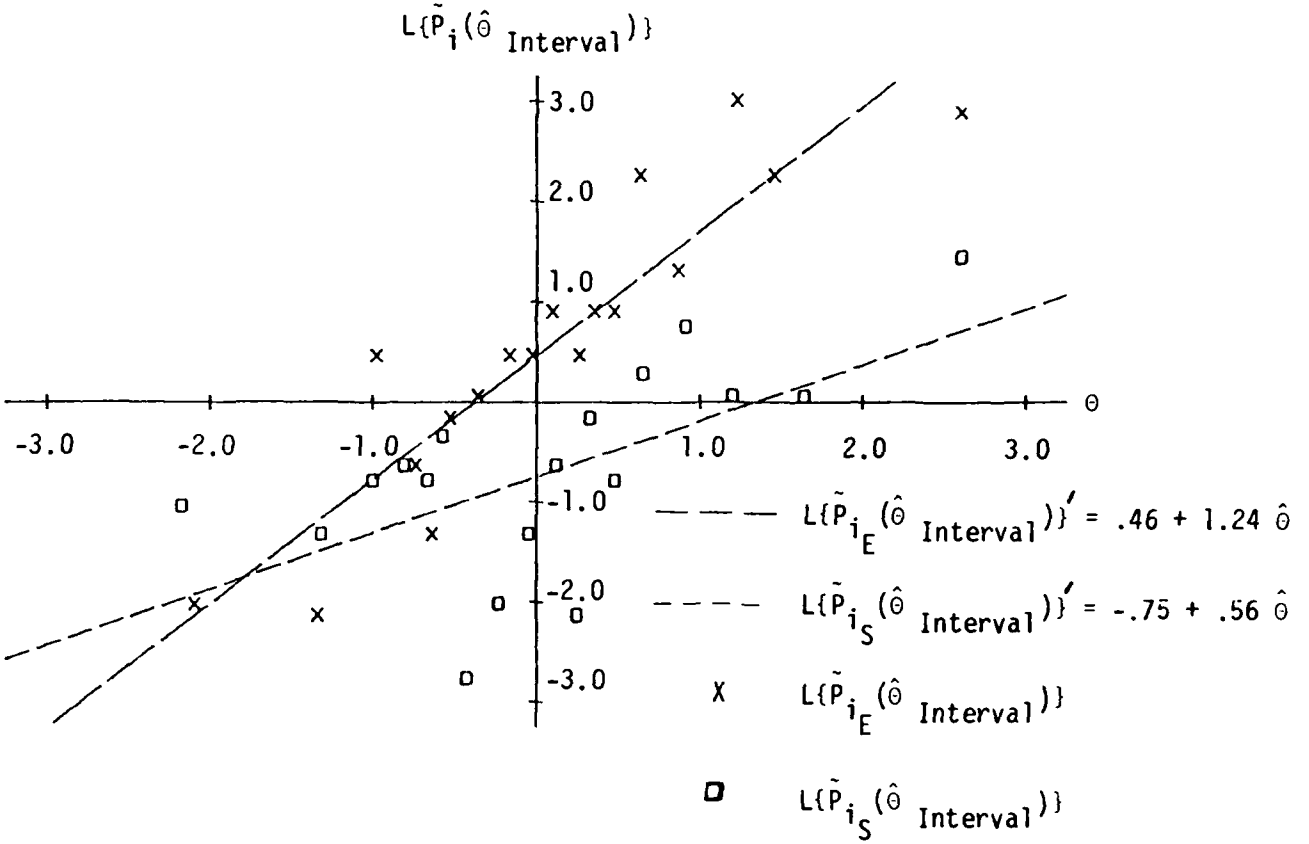
Figure 5. Regression lines and scatter plots for regression of logit transformed proportions of affirmative responses for 18 $\hat{\theta}$ intervals on $\hat{\theta}$ for English and Spanish versions of the JDI item "Influential."

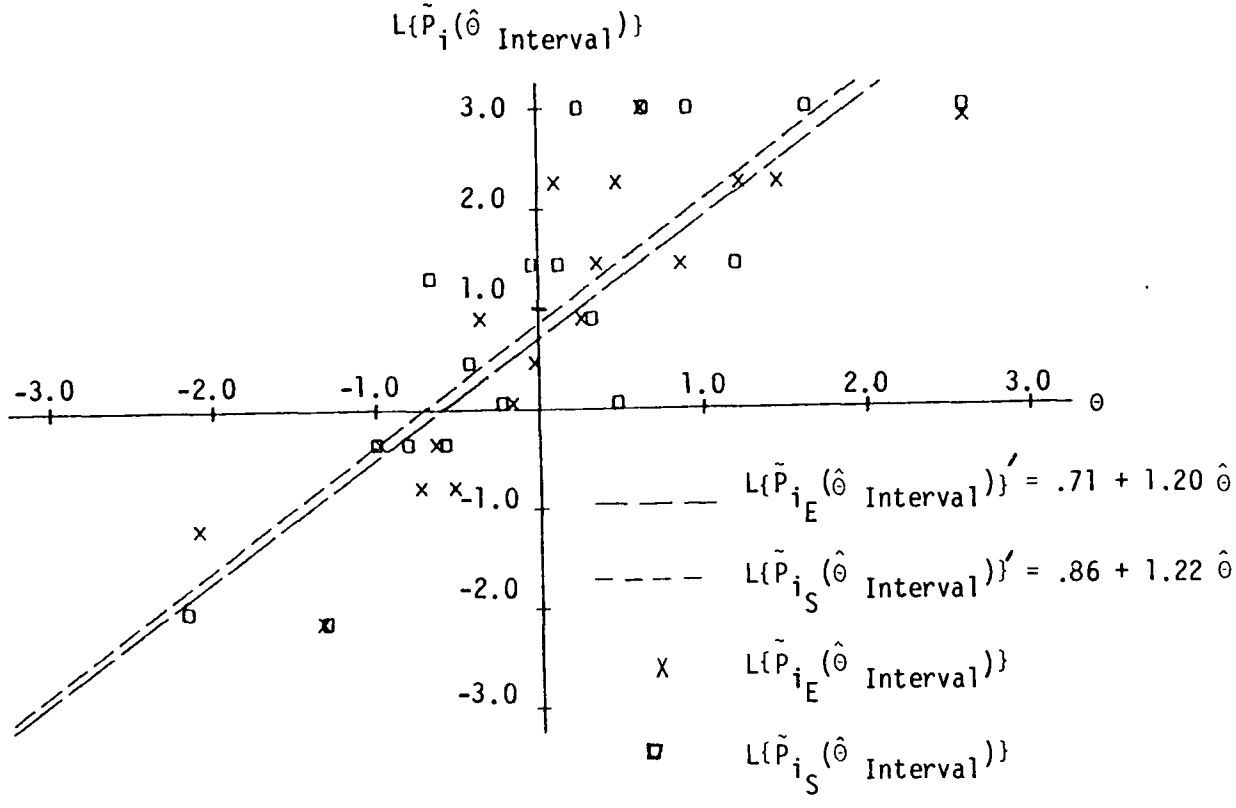

$\hat{\Theta}_S' = -.01 + .96\ \hat{\Theta}_E$
$\hat{\Theta}_S$
4.0
3.0
2.0
1.0
-4.0
-3.0
-2.0
-1.0
1.0
2.0
3.0
4.0
$\hat{\Theta}_E$
-1.0
-2.0
-3.0
-4.0


$\hat{b}_S' = .01 + 1.02\ \hat{b}_E$
$\hat{b}_S$
$\hat{b}_E$
3.0
2.0
1.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
1.0
2.0
3.0

Proportion of Affirmative Responses for $\hat{\theta}$ Interval




$L\{\tilde{P}_i(\hat{\theta}_{Interval})\}$
3.0
2.0
1.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
1.0
2.0
3.0
θ
$L\{\tilde{P}_{i_E}(\hat{\theta}_{Interval})\}' = .46 + 1.24\ \hat{\theta}$
$L\{\tilde{P}_{i_S}(\hat{\theta}_{Interval})\}' = -.75 + .56\ \hat{\theta}$
$L\{\tilde{P}_{i_E}(\hat{\theta}_{Interval})\}$
$L\{\tilde{P}_{i_S}(\hat{\theta}_{Interval})\}$


$L\{\tilde{P}_i(\hat{\theta}\ \text{Interval})\}$
3.0
2.0
1.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
1.0
2.0
3.0
θ
$L\{\tilde{P}_{i_E}(\hat{\theta}\ \text{Interval})\}' = .71 + 1.20\ \hat{\theta}$
$L\{\tilde{P}_{i_S}(\hat{\theta}\ \text{Interval})\}' = .86 + 1.22\ \hat{\theta}$
$L\{\tilde{P}_{i_E}(\hat{\theta}\ \text{Interval})\}$
$L\{\tilde{P}_{i_S}(\hat{\theta}\ \text{Interval})\}$

DISTRIBUTION LIST

Defense Documentation Center
ATTN: DDC-TC
Accessions Division
Cameron Station
Alexandria, VA 22314

Library of Congress
Science and Technology Division
Washington, DC 20540

Chief of Naval Research
Office of Naval Research
Code 452
800 N. Quincy Street
Arlington, VA 22217

Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20375

Commanding Officer
ONR Branch Office
1030 E. Green Street
Pasadena, CA 91106

Psychologist
ONR Branch Office
1030 E. Green Street
Pasadena, CA 91106

Commanding Officer
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

Commanding Officer
ONR Branch Office
Bldg. 114, Section D
666 Summer Street
Boston, MA 02210

Psychologist
ONR Branch Office
Bldg, 114, Section D
666 Summer Street
Boston, MA 02210

Office of Naval Research
Director, Technology Programs
Code 200
800 N. Quincy Street
Arlington, VA 22217

Deputy, Chief of Naval Operations
(Manpower, Personnel, and Training)
Scientific Advisor to DCNO (Op-01T)
2705 Arlington Annex
Washnigton, DC 20350

Deputy Chief of Naval Operations
(Manpower, Personnel, and Training)
Director, Human Resource Management
Division (Op-15)
Department of the Navy
Washington, DC 20350

Deputy Chief of Naval Operations
(Manpower, Personnel, and Training)
Head, Research, Development, and
Studies Branch (Op-102)
1812 Arlington Annex
Washington, DC 30250

Deputy Chief of Naval Operations
(Manpower, Personnel, and Training)
Director, Human Resource Management
Plans and Policy Branch (Op-150)
Department of the Navy
Washington, DC 20350

Chief of Naval Operations
Head, Manpower, Personnel, Training
and Reserves Team (Op-964D)
The Pentagon 4A578
Washington, DC 20350

Chief of Naval Operations
Assistant, Personnel Logistics
Planning (Op-987P10)
The Pentagon, 5D772
Washington, DC 20350

Naval Material Command
Program Administrato, Manpower,
Personnel, and Training
Code 08T244
1044 Crystal Plaza #5
Washington, DC 20360

Naval Material Command
Management Training Center
NMAT 09M32
Jefferson Plaza, BLDG #2, Rm 150
1421 Jefferson Davis Highway
Arlington, VA 20360

Commanding Officer
Naval Personnel R&D Center
San Diego, CA 92152

Navy Personnel R&D Center
Washington Liaison Office
Building 200, 2N
Washington Navy Yard
Washington, DC 20374

Commanding Officer
Naval Heatlh Research Center
San Diego, CA

Commanding Officer
Naval Submarine Medical
Research Laboratory
Naval Submarine Base
New London, Box 900
Groton, CT 06340

Director, Medical Service Corps
Bureau of Medicine and Surgery
Code 23
Department of the Navy
Washington, DC 20372

Naval Aerospace Medical
Research Lab
Naval Air Station
Pensacola, FL 32508

CDR Robert Kennedy
Officer in Charge
Naval Aerospace Medical
Research Laboaratoy Detachment
Box 2940, Michoud Station
New Orleans, LA 70129

National Naval Medical Center
Psychology Department
Bethesda, MD 20014

Commanding Officer
Navy Medical R&D Command
Bethesda, MD 20014

Naval Postgraduate School
ATTN: Dr. Richard S. Elster
Department of Adminstrative Sciences
Monterey, CA 93940

Naval Postgraduate School
ATTN: Professor John Senger
Operations Research and
Adminstrative Science
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Code 1424
Monterey, CA 93940

Officer in Charge
Human Resource Management Detachment
Naval Air Station
Alameda, CA 94591

Officer in Charge
Human Resource Management Detachment
Naval Submarine Base New London
P.O. Box 81
Groton, CT 06340

Officer in Charge
Human Resource Management Division
Naval Air Station
Mayport, FL 32228

Commanding Officer
Human Resource Management Center
Pearl Harbor, HI 96860

Officer in Charge
Human Resource Management Detachment
Naval Base
Charleston, SC 29408

Human Resource Management School
Naval Air Station Memphis (96)
Millington, TN 38054

Commanding Officer
Human Resource Management Center
5621-23 Tidewater Drive
Norfolk, VA 23511

Officer in Charge
Human Resource Management Detachment
Naval Air Station Ehidbey Island
Oak Harbor, WA 98278

Commander in Chief
Human Resource Management Division
U.S. Naval Force Europe
FPO New York, 09510

Officer in Charge
Human Resource Management Detachment
COMNAVFOR JAPAN
FPO Seattle 98762

Chief of Naval Education
and Training (N-5)
ACOS Research and Program
Development
Naval Air Station
Pensacola, FL 32508

Navy Recruiting Command
Head, Research and Analysis Branch
Code 434, Room 8001
801 North Randolph Street
Arlington, VA 22203

Commander in Chief
Human Resource Management Division
U.S. Pacific Fleet
Pearl Harbor, HI 96860

Commanding Officer
Human Resource Management School
Naval Air Station Memphis
Millington, TN 38054

Commanding Officer
Human Resource Management Center
1300 Wilson Boulevard
Arlington, VA 22209

Commander in Chief
Human Resource Management Division
U.S. Atlantic Fleet
Norfolk, VA 23511

Commanding Officer
Human Resource Management Center
Box 23
FPO New York 09510

Officer in Charge
Human Resource Management Detachment
Box 60
FPO San Francisco 96651

Naval Amphibious School
Director, Human Resource
Training Department
Naval Amphibious Base
Little Creek
Norfolk, VA 23521

Naval Military Personnel Command
HRM Department (NMPC-6)
Washington, DC 20350

Chief of Navy Technical Training
ATTN: Dr. Norman Kerr, Code 0161
NAS Memphis (75)
Millington, TN 38054

Naval Training Analysis
and Evaluation Group
Orlando, FL 32813

Naval War College
Management Department
Newport, RI 02940

Headquarters, U.S. Marine Corps
ATTN: Dr. A.L. Slafkosky,
Code RD-1
Washington, DC 20380

Deputy Chief of Staff for
Personnel, Research Office
ATTN: DAPE-PBR
Washington, DC 20310

Army Research Institute
Field Unit - Leavenworth
P.O. Box 3122
Fort Leavenworth, KS 66127

Air University Library
LSE 76-443
Maxwell AFB, AL 36112

Air Force Institute of Technology
AFIT/LSGR (1t. Col. Umstot)
Wright-Patterson AFB
Dayton, OH 45433

AFMPC/DPMYP
(Research and Measurement Division)
Randolph AFB
Universal City, TX 78148

Dr. H. Russell Bernard
Department of Sociology
and Anthropology
West Virginia University
Morgantown, WV 26506

Dr. Michael Borus
Ohio State University
Columbus, OH 43210

Commanding Officer
Naval Training Equipment Center
Orlando, FL 32813

Commandant of the Marine Corps
Headquarters, U.S. Marine Corps
Code MPI-20
Washington, DC 20380

Army Research Institute
Field Unit - Monterey
P.O. Box 5787
Monterey, CA 93940

Headquarters, FORSCOM
ATTN: AFPR-HR
Ft. McPherson, GA 30330

Technical Director
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

AFOSR/NL (Dr. Fregly)
Building 410
Bolling AFB
Washington, DC 20332

Technical Director
AFHRL/ORS
Brooks AFB
San Antonio, TX 78235

Dr. Clayton P. Alderfer
School of Organization & Management
Yale University
New Haven, CT 06520

Dr. Arthur Blaiwes
Human Factors Laboratory, Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Joseph V. Brady
The Johns Hopkins University
School of Medicine
Division of Behavioral Biology
Baltimore, MD 21205

Mr. Frank Clark
ADTECH/Advanced Technology, Inc.
7923 Jones Branch Drive, Suite 500
McLean, VA 22102

Dr. Stuart W. Cook
University of Colorado
Institute of Behavioral Science
Boulder, CO 80309

Mr. Gerald M. Croan
Westinghouse National Issues Center
Suite 1111
2341 Jefferson Davis Highway
Arlington, VA 22202

Dr. Larry Cummings
University of Wisconson-Madison
Graduate School of Business
Center for the Study of Organizational Performance
1155 Observatory Drive
Madison, WI 53706

Dr. John P. French, Jr
University of Michigan
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106

Dr. Paul S. Goodman
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. J. Richard Hackman
School of Organization and Management
Yale University
56 Hillhouse Avenue
New Haven, CT 06520

Dr. Asa G. Hilliard, JR
The Urban Institute for Human Services, Inc.
P.O. Box 15068
San Francisco, CA 94115

Dr. Edna J. Hunter
United States International University
School of Human Behavior
P.O. Box 26110
San Diego, CA 92126

Dr. Rudi Klauss
Syracuse University
Public Administration Department
Maxwell School
Syracuse, NY 13210

Dr. Judi Komaki
Georgia Institute of Technology
Engineering Experiment Station
Atlanta, GA 30332

Dr. Edward E. Lawler
Battelle Human Affiars Research Centers
P.O. Box 5395
4000 N.E., 41st Street
Seattle, WA 98105

Dr. Edwin A. Locke
University of Maryland
College of Business and Management and Department of Psychology
College Park, MD 20742

Dr. Ben Morgan
Performance Assessment Laboratory
Old Dominion University
Norfolk, VA 23508

Dr. Richard T. Mowday
Graduate School of Management and Business
University of Oregon
Eugene, OR 97403

Dr. Joseph Olmstead
Human Resources Research Organization
300 North Washington Street
Alexandria, VA 22314

Dr. Thomas M. Ostrom
The Ohio State University
Department of Psychology
116E Stadium
404C West 17th Avenue
Columbus, OH 43210

Dr. Irwin G. Sarason
University of Washington
Department of Psychology
Seattle, WA 98195

Dr. Saul B. Sells
Texas Christian University
Institute of Behavioral Research
Drawer C
Fort Worth, TX 76129

Dr. Richard Steers
Graduate School of Management
adn Business
University of Oregon
Eugene, OR 97403

Dr. William H. Mobley
University of South Carolina
College of Business Administration
Columbia, SC 29208

Dr. Al. Rhode
Information Spectrum, Inc.
1745 S. Jefferson Davis Highway
Arlington, VA 22202

Dr. Donald Wise
MATHTECH, Inc.
P.O. Box 2392
Princeton, NJ 08540

Dr. George E. Rowland
Temple University, The Merit Center
Ritter Annex, 9th Floor
College of Education
Philadelphia, PA 19122

Dr. Benjamin Schneider
Michigan State University
East Lansing, MI 48824

Dr. H. Wallace Sinaiko
Program Director, Manpower Research
and Advisory Services
Smithsonian Institution
801 N. Pitt Street, Suite 120
Alexandria, VA 22314

Dr. Vincent Carroll
University of Pennsylvania
Warton Applied Research Center
Philadelphia, PA 19104

Dr. Richard Morey
Duke University
Graduate School of Business
Administration
Durham, NC 27706

Dr. Lee Sechrest
Florida State University
Department of Psychology
Tallahassee, FL 32306