An Application of the Continuous Response Level Model to Personality Measurement

Isaac I. Bejar
University of Minnesota

This paper reports an application of Samejima's latent trait model for continuous responses. A brief review of latent trait theory is presented, including an elaboration of the theory for test responses other than dichotomous responses, in order to put the continuous model in perspective. The model is then applied using the *Impulsivity* and *Harmavoidance* scales of Jackson's Personality Research Form. Special attention is given to the requirement that the model be invariant across populations and sex groups. Results showed that responses from males fit the model better than those from females, especially for the Harmavoidance scale. The practical and theoretical implications of the study are discussed.

The fundamental problem of test theory is to infer the true position of testees on a trait or traits of interest, given their responses to a set of test items. In order to make that inference, some assumptions or models are postulated about the relationship of the trait(s) to the responses. A family of these models, together with the concepts surrounding them, is known collectively as *latent trait theory.*

Latent trait theory originated with Ferguson's (1942) and Lawley's (1943) introduction of the normal ogive item characteristic function for dichotomous responses. The theory was further elaborated in the context of attitude measurement by Lazarsfeld (1959), and by Lord (1952), Tucker (1946), and others in the context of ability measurement. The early results and some new developments by Birnbaum were incorporated into a textbook by Lord and Novick (1968). Rasch's (1960) work is also part of latent trait theory but has been developed independently from the mainstream of the work being done in the United States.

Up to and including Lord and Novick's book, the theoretical developments were oriented exclusively toward dichotomous scoring. Samejima (1969, 1972, 1973) and Bock (1972) have extended the theory to the polychotomous case, where responses are categorized into more than two categories, as well as to the continuous case. Additionally, Samejima has generalized the theory to the multidimensional case (Samejima, 1972*a*, 1974*b*). The extension of the theory to the polychotomous and continuous cases is important from an applied perspective, since both cases allow a more precise estimation of a person's position on the latent trait.

This paper reports what appears to be the first application of Samejima's model for continuous responses. Before turning to that application, however, some basic concepts underlying latent trait theory will be reviewed.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 4 Fall 1977 pp. 509–521
© Copyright 1977 West Publishing Co.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Latent Trait Theory

Response Levels

Latent trait theory characterizes testees' trait levels by their position on a continuum, denoted by Θ, which is assumed to be $-\infty < \Theta < \infty$. The presentation of the g^{th} test item to a testee of trait level Θ elicits a latent response, γ_g. For a variety of reasons which may collectively be called *error,* the relationship between γ_g and Θ is not a perfect one, but it is assumed to be linear, i.e.,

$$\gamma_g = \rho_g \theta + e_g \qquad [1]$$

where ϱ_g is the correlation of γ_g and Θ, and e_g is an error component which is assumed to be independent of and not correlated with any other e_g and Θ. Although γ_g and Θ are continuous, the response that is actually recorded need not be. Several response levels can be distinguished, depending on the grossness of the observed response. If observed responses fall into one of two categories (e.g., correct-incorrect, true-false), the responses are at the *dichotomous response level.* If there are more than two, but a finite number of categories, the responses can be considered to be at the *polychotomous or graded response level* (e.g., response to a mathematics problem may be scored in terms of degree of correctness). Finally, if the number of response categories is infinitely large, responses can be considered to be at the *continuous response level.*

Central to latent trait theory is the formulation of models which give the probability of occurrence of the *observed* response. To derive some specific models, distributional assumptions about e_g and γ_g must be introduced. In particular, the distribution of γ_g, given Θ, is characterized solely by the distribution of e_g (Samejima, 1974*a*). Assuming e_g distributed normally with a mean of zero and a standard deviation of 1 (as is done here), leads to the normal ogive model.

More concretely:

1. In the dichotomous response level, the probability of passing or endorsing the g^{th} item for a fixed value of Θ is given for all items by the *item characteristic function* of the normal ogive form,

$$P_g(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta - b_g)} e^{t^2/2} dt \qquad [2]$$

where $t = (\gamma_g - \rho_g \theta)/\sigma_g$, $\sigma_g = \sqrt{1-\rho_g^2}$

and $\varrho_g\Theta$ and σ_g represent the mean and standard deviation of γ_g conditional on Θ.

2. In the polychotomous response level, the probability of scoring in category x_g within the g^{th} item for a fixed value of Θ is given for all items by the *operating characteristic function* (Samejima, 1969) of the normal ogive form,

$$P_{x_g}(\theta) = \frac{1}{\sqrt{2\pi}} \int_{a_g(\theta - b_{x_{g+1}})}^{a_g(\theta - b_{x_g})} e^{-t^2/2} dt . \qquad [3]$$

3. In the continuous response level, the probability of responding with z_g to the g^{th} item for a fixed value of Θ is given by the *operating density characteristic function* (Samejima, 1973) of the normal ogive form,

$$H_{z_g}(\theta) = \frac{a_g}{\sqrt{2\pi}} \left[e^{-a_g^2 (\theta - b_{z_g})^2/2} \right] \left[\frac{d\, b_{z_g}}{dz_g} \right] . \qquad [4]$$

It should be noted that z_g is assumed to be $0 < z_g < 1$.

Then, by virtue of local independence (see Lord & Novick, 1968, chap. 16), it follows that the (unconditional) distribution of the vector of latent responses $\gamma = [\gamma_g]$, $g = 1, 2 \ldots n$, is multivariate normal with mean vector 0 and covariance matrix

$$\Sigma = \Lambda\Lambda' + \Psi \qquad [5]$$

where $\Lambda = [\varrho_g]$, $g = 1, 2 \ldots n$, and Ψ is a diagonal matrix with elements $[1 - \varrho_g^2]$, $g = 1, 2 \ldots n$.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Interpretation and Estimation

Each of Equations 2 through 4 is seen to depend on Θ, a_g and either b_g, b_{x_g} or b_{z_g}, depending on the response level. In each case, a_g has the same interpretation; it is an index of the discriminating power of the item and is a function of $\hat{\varrho}_g$, the estimated item-trait correlation. Under the multinormality and unidimensionality assumptions, a_g is estimated by

$$a_g = \frac{\hat{\rho}_g}{\sqrt{1-\hat{\rho}_g^2}} \quad . \qquad [6]$$

b_g and b_{x_g} apply to the dichotomous and graded response level, respectively, and are considered difficulty parameters since they are related to the proportion of testees who pass the item or respond in a given category. Under the assumptions of multinormality and unidimensionality, these parameters are estimated by

$$b_g = \Phi^{-1}(p_g)/\hat{\rho}_g \qquad [7]$$

and

$$b_{x_g} = \Phi^{-1}(p_{x_g})/\hat{\rho}_g \qquad [8]$$

where p_g is the proportion passing the item, and p_{x_g} is the proportion scoring in category x_g and above.[1]

In the continuous case we have,

$$b_{z_g} = g^{-1}(z_g)/\hat{\rho}_g \qquad [9]$$

where g^{-1} is a function that transforms each value of z_g into a normal deviate. In the special case where z_g is uniformly distributed,

$$b_{z_g} = \Phi^{-1}(z_g)/\hat{\rho}_g \quad . \qquad [10]$$

Item parameter estimation. The fit of the data to the normal ogive model has relied mainly on whether the factor analysis of the inter-item correlation matrix yields a single factor (e.g., Indow & Samejima, 1966). The correlation matrix consists, in principle, of tetrachoric and polychoric correlations in the dichotomous and polychotomous case respcetively.[2] In the continuous case, the inter-item correlation matrix consists of product-moment correlations computed on the z_g's transformed by g^{-1}. If the hypothesis of a single factor is accepted, then the item loadings on that single factor are taken as the $\hat{\varrho}_g$'s, which in turn are used in estimating b_g, b_{xg}, and b_{z_g}, depending on the response level, as well as a_g.

The procedure described in the previous paragraph is what Bock and Lieberman (1970) call the heuristic estimation procedure. In the same paper, they developed what they called the unconditional maximum likelihood procedure. Although this latter procedure is preferable from a statistical point of view, its application with more than 10 to 12 items is not practical. Moreover, Bock and Lieberman noted very good agreement between the heuristic and unconditional maximum likelihood solution.

The use of factor analysis in the estimation of normal ogive parameters is not only convenient, but as Samejima (1974*a*) has noted, it has some formal similarities in the normal ogive and linear factor analysis models. In particular, she noted that a sufficient condition that leads to the complete latent space is factorial invariance over all populations of interest. That is, the structure of the j^{th} population is given by

$$\Sigma_j = \Lambda\Phi_j\Lambda' + \Psi \qquad [11]$$

where Σ_j is the j^{th} population covariance matrix, Λ is a factor pattern matrix which is the same for all populations, Φ_j is the inter-trait covariance matrix for the j^{th} population (i.e., the relationship among the traits need not be invariant over populations), and Ψ is a diagonal matrix of unique variances which is also common to all

[1] The response categories are labeled *0, 1, 2 . . . m_g, m_g+ 1. 0* and $m_g + 1$ are, in a sense, below and above the realm of possibility and $b_{x_{g=0}} = -\infty$ and $b_{x_{g=m_g+1}} = \infty$.

[2] In practice, however, such a matrix may not be Grammian as required by factor analysis, but recently Christofferson (1975) has devised a procedure which avoids that problem in the dichotomous case.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

populations. In the single-factor case, Equation 11 reduces to

$$\Sigma_j = \Lambda\Lambda' + \Psi \qquad [12]$$

with Λ and Ψ as described earlier.

Estimating Θ. The method of estimation to be discussed here is the *maximum likelihood method.* Let the responses to a set of items be denoted by the vector $\mathbf{v}$, the elements of which are *1*'s and *0*'s in the dichotomous case; and x_g's and z_g's in the graded and continuous cases. For convenience, assume that all elements in $\mathbf{v}$ are legitimate, e.g., no missing responses are included. The likelihood function of $\mathbf{v}$ is, by virtue of the principle of local independence, the product of the individual item likelihoods. This is,

$$L_v(\theta) = \prod_{g=1}^{n} L_g(\theta) \qquad [13]$$

where $L_g(\Theta)$ is the likelihood function for the responses to the g^{th} item. Thus, in the dichotomous case

$$L_v(\theta) = \prod_{g=1}^{n} P_g(\theta)^{u_g} Q_g(\theta)^{1-u_g} \qquad [14]$$

where $P_g(\Theta)$ is given by Equation 2, $Q_g(\Theta) = 1 - P_g(\Theta)$, and u_g is *1* or *0* depending on whether the response is correct or incorrect.

In the polychotomous case the likelihood function is given by

$$L_v(\theta) = \prod_{g=1}^{n} P_{x_g}(\theta) \qquad [15]$$

where $P_{x_g}(\Theta)$ is given by Equation 3. Finally, in the continuous case

$$L_v(\theta) = \prod_{g=1}^{n} H_{z_g}(\theta) \qquad [16]$$

where $H_{z_g}(\Theta)$ is given by Equation 4.

According to the likelihood principle, the maximum likelihood estimate is that value of Θ which maximizes $L_v(\Theta)$. That value can, in principle, be obtained by differentiating $L_v(\Theta)$ with respect to Θ, equating to zero, and solving for Θ; that is,

$$\left[\frac{\partial L_v(\theta)}{\partial \theta}\right]_{\theta=\hat{\theta}} = 0 \qquad [17]$$

$$\left[\frac{\partial \log L_v(\theta)}{\partial \theta}\right]_{\hat{\theta}=\theta} = 0 \quad .$$

It is usually more convenient, however, to work with the log of the likelihood, i.e.,

$$\left[\sum_g A_g(\theta)\right]_{\theta=\hat{\theta}} = 0 \qquad [18]$$

where $\partial \log L_g(\Theta)/\partial\Theta = A_g(\Theta)$. Samejima (1969) has called $A_g(\Theta)$ the *basic function* since many properties of a particular model can be derived from it.

It should be noted that in the dichotomous and graded cases, the solution of Equation 18 does not yield an explicit formula for Θ, and therefore, numerical methods are required to find Θ. In contrast, in the continuous case $A_g(\Theta) = -a_g^2(\Theta - b_{z_g})$; and substituting into Equation 18 gives

$$0 = \sum_{g=1}^{n} -a_g^2(\theta - b_{z_g}) \qquad [19]$$

$$= \sum_{g=1}^{n} -a_g^2\theta + \sum_{g=1}^{n} a_g^2 b_{z_g}$$

and solving for Θ,

$$\hat{\theta} = \frac{\sum_{g=1}^{n} a_g^2 b_{z_g}}{\sum_{g=1}^{n} a_g^2} \qquad [20]$$

which is the maximum likelihood of Θ in the continuous case. Thus, given a set of responses

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

z_g ($g = 1, 2 \ldots n$) to estimate Θ, b_{z_g} is computed for each item according to Equation 9, weighted by a_g^2, summed over all items and finally divided by Σa_g^2.

Efficiency and Information

Efficiency. The efficiency of an estimator T of Θ, denoted $T(\Theta)$, is defined, using the present notation (see Lindgren, 1968, p. 275), as

$$e[T(\theta)] = \frac{\left\{\mathrm{COV}\left[\sum_g A_g(\theta), T(\theta)\right]\right\}^2}{\mathrm{VAR}\left[\sum_g A_g(\theta)\right] \mathrm{VAR}\left[T(\theta)\right]} \qquad [21]$$

that is, the squared correlation of $\Sigma A_g(\Theta)$ and $T(\Theta)$. $A(\Theta)$ is the basic function defined previously. It can be shown (Lindgren, 1968, p. 273) that the expectation of $A_g(\Theta)$ is zero and therefore its variance is simply

$$\mathrm{VAR}\left[\sum_g A_g(\theta)\right] = E\left\{\left[\sum_g A_g(\theta)\right]^2\right\} \qquad [22]$$

$$= \left\{\left[\frac{\partial \log L_v(\theta)}{\partial \theta}\right]\right\} .$$

Also, when T is an unbiased estimator, the numerator of Equation 21 is 1.0.

An estimator is said to be efficient if $e[T(\Theta)] = 1$. This implies that the variance of the estimator, VAR$[T(\Theta)]$, is the reciprocal of the right side of Equation 22; that is,

$$\mathrm{VAR}[T(\theta)] = \frac{1}{\mathrm{VAR}\left[\sum_g A_g(\theta)\right]} \qquad [23]$$

$$= \frac{1}{E\left\{\left[\frac{\partial \log L_v(\theta)}{\partial \theta}\right]^2\right\}}$$

as can readily be seen by substitution into Equation 22. Since maximum likelihood estimators do, in fact, have variance asymptotically equal to the right-hand side of Equation 23 (Kendall & Stuart, 1961, chap. 18, pp. 43-44), maximum likelihood estimates are asymptotically efficient. Also, the right-hand side of Equation 23 is the minimum attainable variance (Kendall & Stuart, 1961, chap. 17, pp. 8-10); and, in this sense, maximum likelihood estimators are asymptotically best.

Information. Samejima (1969) has defined the amount of information in a test as

$$I(\theta) = E\left\{\left[\frac{\partial \log L_v(\theta)}{\partial \theta}\right]^2\right\} \qquad [24]$$

i.e., the reciprocal of VAR $[T(\Theta)]$. Since VAR $[T(\Theta)]$ is the minimum attainable variance, $I(\Theta)$ is, in a sense, a measure of "the best that can be done" in estimating Θ, assuming a particular model. Because maximum likelihood estimates are asymptotically minimum variance estimates, they are also asymptotically maximally informative.

The concepts of information and efficiency are far more useful than their (squared) classical test theory counterparts — reliability and error of measurement. At least three advantages are evident. First, information and efficiency are population-free concepts and do not depend on the variability of the trait in different groups, as do reliability and error of measurement. Second, unlike error of measurement, which is an overall index, $I(\Theta)$ is a function of Θ. This means that the precision of a testing procedure can be expressed at different levels of ability. Finally, through the use of information functions, testing procedures measuring the same construct with possibly different items can be compared. This is true even though $I(\Theta)$ is not invariant over transformations of Θ (Lord, 1974), since changing the metric of Θ alters the shape of $I(\Theta)$; but the ratio of the information functions is invariant over such transformations. This ratio, which can be called *relative information*, may be denoted as

$$RI_{A/B} = I_A(\theta)/I_B(\theta) \qquad [25]$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

where $I_A(\Theta)$ is the information function of testing procedure A, and $I_B(\Theta)$ is the information function of testing procedure B. Its interpretation is in terms of test length: if $RI_{A/B} > 1.00$ at a given value of Θ, this means that Test A measures as well as Test B lengthened by a factor of $RI_{A/B}$.

The relative information of the dichotomous, polychotomous, and continuous response levels can be compared by computing Equation 24 at several points of Θ for the corresponding models and then computing Equation 25. Bejar (1975) has performed such comparisons for the dichotomous, polychotomous, and continuous response levels and noted that substantial gains may result. Moreover, in the continuous case, the information function is not only higher than those of the polychotomous and dichotomous cases, but, additionally, it is constant across all levels of Θ. This means that unlike the dichotomous and polychotomous response levels, in the continuous case testees at all levels of the trait are measured with equal precision.

An Application of the Continuous Response Model

The continuous response level has been discussed (de Finetti, 1965; Shuford, Albert, & Massengill, 1966) and applied (see Echternacht, 1972) previously in the context of multiple-choice ability tests under the rubric of "probabilistic testing." The rationale behind this approach is that knowledge can be quantified in terms of the testee's subjective probability about the correctness of each alternative. To insure an honest and careful response, scoring formulas are developed with the property of maximizing the testee's subjective expected score.

Research to this date has not established conclusively the practical advantages of probabilistic testing. Perhaps part of the problem is that testees are not able to translate their knowledge into a subjective probability statement with the limited practice they are usually allowed. This suggests that a more fruitful application of the continuous response level may be with naturally continuous responses, e.g., response latencies. In fact, according to Bock (1973):

> [In the spatial tests] . . . the key to improved testing may be in exploiting information in the response latencies. Testing technology has by now reached a stage where measurement of response times could be routinely incorporated in the data collection, even in group testing, and the new methods of latent trait estimation could be extended to recover information from latencies as well as from the alternatives chosen (p. 455-456).

Personality measurement appears to be another natural application of the continuous response level model. The feasibility of such an application was investigated in the present study.

Method

Subjects. A total of 350 volunteer students, enrolled in the introductory psychology course at the University of Minnesota, participated in the study. Of these, 29 were eliminated for various reasons, primarily incomplete responses. Of the remaining students, 178 were males and 143 were females.

Tests and instructions. The *Form A* Impulsivity (*Im*) and Harmavoidance (*Ha*) scales from the Jackson Personality Research Form (*PRF*; Jackson, 1967) were selected for study. Each of these scales consists of 20 items with 10 true and 10 false items. (The scales are presumably unidimensional because of the way they were constructed.) According to the manual, a high scorer on the *Ha* scale "does not enjoy exciting activities, especially if danger is involved; avoids risk of bodily harm; seeks to maximize personal safety." The high scorer on the *Im* scale "tends to act on the 'spur of the moment' and without deliberation; gives vent readily to feeling and wishes; speaks freely; may be volatile in emotional expression."

The wording of the items was retained as it appeared in the original testing booklet. Normally, testees respond to the *PRF* by stating whether an item is true or false. The instructions were modified for the present investigation by instructing testees to express their degree of

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

agreement on a scale from *0* to *100*, where responses close to *0* indicated little agreement and responses close to *100* indicated a high degree of agreement. Students were encouraged to use the entire continuum in order to reduce stereotypic responses. For analysis purposes, the responses were divided by 100 to conform with the restriction that the response be greater than zero and less than 1.

Analysis

As stated earlier, a sufficient condition for the adequacy of the model is the invariance of the model parameters with respect to populations of interest. Here, the populations consisted of the male and female subgroups. Populations of interest may be defined by socioeconomic level or other demographic variables in addition to, or instead of, sex; the psychometrician must decide what populations are of interest.

To analyze the data, responses to false items were reversed; i.e., if a response to a false item was z^*, it was recoded to $z = 1 - z^*$. This recoding has the effect of orienting all the items in the same direction. Responses to the g^{th} item were transformed to latent responses by the formula

$$\hat{\gamma}_{ig} = \Phi^{-1}[(P_{ig} - .5)/N] \qquad [26]$$

where P_{ig} is the response given by the i^{th} subject to the g^{th} item, Φ^{-1} is the inverse normal function and N is the sum of males and females (321). Upon transformation, $\hat{\gamma}_g$ is approximately normally distributed with a mean of zero and standard deviation of 1 over both sexes; but neither the mean, the standard deviation, nor the form of distribution need to be identical for each sex upon transformation. The first two are not required by the model; in other words, the latent responses of males and females may have different locations and dispersions, as does the trait itself. However, the form of the distribution should be the same, i.e., normal, within each sex.

Thus, one of the criteria of the feasibility of this application of the model was the homogeneity of distributional form between the sexes. The second criterion was factorial invariance over sexes; that is, the loadings of each item on the single factor as well as the unique loadings should be the same for both sexes. The first criterion implies that g^{-1}, the function that maps z into Θ, is the same for males and females. The second criterion implies that the inter-item covariance matrices for the two sexes are accounted for by a single factor with the same composition in each sex. If these two criteria are satisfied, the sex of a respondent need not be identified to estimate his/her position on the latent trait. Moreover, the expected response to an item—e.g., $E(z_g\Theta)$—is the same for any given Θ regardless of sex, and in that sense the test is "bias free" with respect to sex.

To examine the homogeneity of distributional form, the $\hat{\gamma}_g$'s for each item were tested for normality in each sex separately using the Kolmogorov-Smirnov statistic (e.g., Lindgren, 1968). According to this test, if the observed cumulative frequency exceeds the theoretically expected frequency at any point, the null hypothesis of equality of distribution is rejected. More specifically, the statistic may be written

$$D_g = \max \left| F_o(z_g) - F_e(z_g) \right| \qquad [27]$$

where $F_o(z_g)$ is the observed cumulative distribution for item g, $F_e(z_g)$ is the expected cumulative distribution, and D_g denotes the largest absolute difference between the two distributions. If that difference exceeds a critical value, the null hypothesis is rejected.

Results

In the present case, $F_e(z_g)$ was specified to be normal; and the critical values were .11 and .10 for females and males, respectively. All the *Im* items were judged to be normal for males and females. Four of the 20 *Ha* items—the eighth, seventeenth, eighteenth, and twentieth—had observed D_g's of .11, .12, .12, and .14, respectively, for males. Thus, the assumption of homogeneity of distribution by sex was not fulfilled by those four items. In general, however, the assumption

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

seems reasonable, especially since some "significant results" are expected to occur by chance.

The first step taken in examining the factorial invariance question was to run a maximum likelihood factor analysis separately for the two sex groups, specifying a single factor. The resulting chi-square statistics for *Im* were 350 and 366, for females and males, respectively. For the *Ha* items, the corresponding chi-squares were 357 and 301. In all four cases, the degrees of freedom were 170 since all 20 items were included in the analysis; the associated *p*-values were below .0001, suggesting the implausibility of a single factor.

To detect the relative contribution of an item to the lack of fit, the factor analysis was re-run without that item. The resulting chi-squares are seen in Table 1. A reasonable rule of thumb for assessing the relative contribution of an item is to compute the difference in chi-squares for the solution with all 20 items minus the chi-square for the solution with that item removed. For example, for the females' *Im* data, the difference for item 1 is 350 − 312 = 38, whereas that for item 6 is 350 − 330 = 20. Hence, item 1 has a poorer fit with respect to the unidimensionality hypothesis than item 6. Examining these differences, seven *Ha* and seven *Im* items were chosen for further analysis because of their relatively better fit for both males and females. Having found this core of presumably unidimensional items, the data were analyzed simultaneously for

Table 1
Chi-Square Values After Removing One Item for the *Im* and *Ha* Scales

		Scale			
		Impulsivity		Harm-avoidance	
Item Removed	df	Females	Males	Females	Males
none	170	350	366	357	301
1	152	312	315	329	263
2	152	317	313	323	273
3	152	308	329	313	270
4	152	303	331	321	267
5	152	310	329	327	271
6	152	330	324	302	269
7	152	315	317	319	245
8	152	311	331	314	278
9	152	316	345	314	271
10	152	320	339	319	278
11	152	323	335	323	268
12	152	314	328	315	280
13	152	323	340	304	262
14	152	292	330	315	290
15	152	310	308	328	278
16	152	321	335	315	253
17	152	329	324	325	281
18	152	316	330	330	264
19	152	302	336	316	270
20	152	301	303	327	250

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

males and females to test invariance and estimate item-trait correlations (i.e., loadings) from which the item parameters could then be estimated. The program *SIFASP* (van Thillo & Joreskög, 1970) was used for this purpose.

The variance-covariance matrices for males and females were submitted for analysis. The program was instructed to scale the male and female matrices so that their weighted average was a correlation matrix. Specifically, each variance-covariance matrix was pre- and post-multiplied by the diagonal matrix $\mathbf{D} = (\text{diag } \mathbf{S})^{-1/2}$ where $\mathbf{S} = (N_m\mathbf{S}_m + N_f\mathbf{S}_f)/(N_m + N_f)$, $\mathbf{S}$ is a covariance matrix, N represents sample size, and the subscripts m and f indicate male and female. In addition, the following restrictions were imposed on the solution: a single factor was to be extracted, and the loadings on that factor and unique variances were to be identical for both males and females. The maximum likelihood solution was scaled so that the weighted average variance of the male and female factor was one.

Table 2
Simultaneous Maximum Likelihood Solution for *Im* and *Hα* Scales

Im Items

Item	Λ	$\psi^{\frac{1}{2}}$	Summary Data
8 (F)	.174	.985	$\chi^2 = 68.3$, df = 40
9 (T)	.334	.942	p = .003
10 (F)	.445	.896	$\phi_f = 1.08$, $\phi_m = .94$
11 (T)	.247	.969	Mean female residual: .08
13 (T)	.502	.865	Mean male residual: .06
16 (F)	.205	.979	
18 (F)	.195	.981	

Hα Items

Item	Λ	$\psi^{\frac{1}{2}}$	Summary Data
2 (F)	.368	.930	$\chi^2 = 65.6$, df = 40
4 (F)	.452	.892	p = .007
5 (T)	.510	.860	$\phi_f = 1.07$, $\phi_m = .95$
10 (F)	.407	.913	Mean female residual: .06
11 (T)	.408	.913	Mean male residual: .08
15 (T)	.502	.865	
19 (T)	.458	.889	

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Scaling of the maximum likelihood solution in this manner in conjunction with the scaling of the variance-covariance matrices yields results which are interpreted as if they were derived from the usual correlation matrix. These results are seen in Table 2.

The first column of Table 2 refers to the number of the items included in the analysis. The *T* and the *F* which appear in parentheses next to each item indicate whether the item is keyed as true or false. The columns labeled Λ and $\Psi^{1/2}$ contain the maximum likelihood estimates of item loadings and item uniquenesses resulting from the simultaneous analysis of males and females.

The statistical fit of the solutions can be evaluated from the chi-square statistic, which was 68.3 and 65.6 for *Im* and *Ha* items, respectively. With 40 degrees of freedom, the probability of the null hypothesis is rather low in both cases. However, the mean absolute value of the residuals, i.e., the difference between the observed and reproduced covariance matrices, was not very large (as can be seen in Table 2) nor did the residuals exhibit any perceptible pattern.

To further examine sex differences, the data were re-analyzed separately for males and females, imposing only the restriction of a single factor. The results in terms of the chi-square statistic are seen in Table 3. Evidently, the unidimensionality assumption was reasonable for the male data but less so for the female data, especially in the case of *Im* items.

Table 3 also reports two estimates of reliability, α and *mlα*. α is the usual estimate, while *mlα* is estimated from the formula

$$\text{ml}\alpha = \frac{(\sum_g \hat{\rho}_g)^2}{(\sum_g \hat{\rho}_g)^2 + \sum_g \psi} \qquad [28]$$

(see Werts, Linn, & Joreskög, 1974), where ψ is the g^{th} item unique variance, corresponding to the diagonal elements of Ψ. For the *Ha* data, α and *mlα* are essentially the same whether estimated from males, females, or males and fe-

Table 3
Comparison of the Male, Female, and Male-Female Solutions

	Scale	
Group	*Im*	*Hα*
Female	$\chi^2 = 33.1$, df = 14, $p < .005$ $\alpha = .367$, $\text{ml}\alpha = .321$	$\chi^2 = 26.2$, df = 14, $p < .02$ $\alpha = .624$, $\text{ml}\alpha = .628$
Male	$\chi^2 = 21.7$, df = 14, $p < .10$ $\alpha = .452$, $\text{ml}\alpha = .448$	$\chi^2 = 20.0$, df = 14 $p < .20$ $\alpha = .638$, $\text{ml}\alpha = .635$
Male-female*	χ^2 68.3, df = 40, $p < .004$ $\text{ml}\alpha = .413$	$\chi^2 = 65.6$, df = 40, $p < .008$ $\text{ml}\alpha = .632$

*from Table 2

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

males. Such is not the case, however, for the *Im* data.

Once a solution is accepted, a_g's and b_{z_g}'s can be computed. The discrimination parameter, a_g, is computed from $\varrho_g \psi_g^{-1/2}$, which is the generalization of Equation 6 needed whenever ϱ_g is estimated simultaneously in two or more populations. In the continuous case, b_{z_g} is a function which is empirically computed; that is, a function from z_g to $\hat{\gamma}_g/\hat{\varrho}_g$ is fitted. Samejima (1973) recommends

$$b_{z_g} = \frac{\log z_g - \log(1-z_g) + c}{d} . \quad [29]$$

Upon obtaining a_g and b_{z_g} for the items, the test may be said to be calibrated, and it is ready for further research.

Discussion

Although the *PRF* was constructed using very stringent psychometric criteria, sex-invariance was not one of them. In view of this fact, it is not surprising that the fit of the model was less than perfect; but it seems likely that better sex-invariance is obtainable, at least for the *Ha* scale, with additional item selection. Nevertheless, even though the statistical fit was not very high, other evidence indicated that the model was reasonable for *Ha* data. For example, the magnitude of the residuals was not large, nor was there a particular pattern among the residuals. Furthermore, α and $ml\alpha$ were almost identical within the male, female, and simultaneous solutions, which suggests the feasibility of unidimensionality in both males and females. Moreover, $ml\alpha$ and α were also very close across the male, female and simultaneous solutions, suggesting the invariance across sex of the item-trait correlation of the *Ha* items. The model, however, does not seem to provide a good representation for the *Im* data. The $ml\alpha$ and α estimates were not consistent within solutions, and, more importantly, unidimensionality was not a reasonable assumption for the females even though it was for the males.

The question may arise as to whether the approach to test construction illustrated here is worth the effort. There are several reasons why it is likely to be. In classical theory, the overriding concern is with the construction of parallel tests, whereas in latent trait theory the concern is with tests which are invariant across populations. Population-invariant tests assume that testees with the same trait levels will respond in a like manner and will thus be estimated to have the same trait levels regardless of the population to which they belong. Such characteristics have important practical implications, including the development of culture-fair tests (e.g., Pine & Weiss, 1976).

On the substantive side, this approach to test construction examines the dimensionality of responses to the same items by different populations. In the example presented here, the responses of females to *Im* items were decidedly multidimensional. This could occur if females do in fact respond to *Im* items as a function of two or more personality traits, in which case this finding would be of substantive interest. Or, it could be a result of some other violation of the local-independence assumption, i.e., Ψ is not a diagonal matrix. If, for example, females were responding in part as a function of the wording of the items—whether they were worded positively or negatively—this would violate the assumption of local independence and introduce correlation among the items which would not be due to the *Im* trait.

Turning to the use of the continuous response level itself, it is not argued here that it is always necessary; rather, psychologists must decide what degree of accuracy is required in a particular application. However, whenever the highest precision of measurement is required, the continuous response level is called for since it not only yields more precise trait estimates at *all* levels of Θ, but in addition, that level of precision remains constant at all trait levels (Bejar, 1975; Samejima, 1973). This contrasts with the dichotomous and polychotomous levels which yield information functions with a peak at some value of Θ but yield increasingly less information for levels of Θ away from that value.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

The internal consistency of the *Ha* data with seven items was about .63. Stepping-up this by a factor of 2.85 to make it comparable to the estimates based on 20 items results in a value of .82, which is comparable to the .80 to .83 estimate reported by Jackson for true-false data. Thus, using reliability as the criterion, there has been no gain in employing the continuous response level. In terms of information, however, different conclusions result.

In the continuous case, $I(\Theta) = \Sigma a_g^2$ (Samejima, 1973). That is, information is constant at all levels of Θ and equal to the sum of the squared item discrimination parameters. In the dichotomous case, $I(\Theta)$ is a curve with a maximum at some Θ, say Θ^*. If all the items within the test are equivalent, the maximum is given by (cf. Birnbaum 1968, p. 462)

$$\Sigma M_g = \Sigma 2 a_g^2 \pi \quad . \qquad [30]$$

If the items are not equivalent, Equation 30 is just an approximation. Now, assuming that the a_g's would have been the same under true-false administration (which seems reasonable since the reliability estimates reported here were close to those reported by Jackson), the ratio $\Sigma a_g^2 / \Sigma M_g \cong 1.57$ gives an indication of the gain in precision due to the use of the continuous response level. In this case, the continuous response level is 57% more informative than the dichotomous level at Θ^*. For values of Θ away from Θ^*, the relative information of the continuous over the dichotomous response level becomes increasingly larger. Consequently, use of the continuous response level with the *Ha* scale results in considerably more precise measurement throughout the trait range compared to the dichotomous responses normally used with the *PRF*.

References

Bejar, I. I. *The relative information of the dichotomous, graded and continuous response levels as a function of item discrimination and difficulty.* Paper presented at the joint meeting of the Psychometric and Classification Society, April 24–26, 1975, Iowa City, IA.

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

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika,* 1972, *37,* 29–52.

Bock, R. D. Word and Image: Sources of the verbal and spatial factors in mental test scores. *Psychometrika,* 1973, *38,* 437–458.

Bock, R.D., & Lieberman, M. L. Fitting a response model for *n* dichotomously scored items. *Psychometrika,* 1970, *35,* 179–197.

Christofferson, A. Factor analysis of dichotomized variables. *Psychometrika,* 1975, *40,* 5–32.

de Finetti, B. Methods of discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology,* 1965, *13,* 87–123.

Echternacht, G. J. The use of confidence testing in objective tests. *Review of Educational Research,* 1972, *42,* 217–235.

Ferguson, G. A. Item selection by the constant process. *Psychometrika,* 1942, *7,* 19–29.

Indow, T., & Samejima, F. *On the results obtained by the absolute scaling model and the Lord model in the field of intelligence.* Yokohama: Psychological Laboratory, Hiyoshi Campus, Keio University, 1966.

Jackson, D. N. *Personality Research Form Manual.* Goshen, NY: Research Psychologists Press, 1967.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (Vol. II). New York: Hafner, 1961.

Lawley, D. N. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh,* 1943, *61,* 273–287.

Lazarsfeld, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 3). New York: McGraw-Hill, 1959.

Lindgren, B. W. *Statistical theory.* New York: MacMillan, 1968.

Lord, F. M. A theory of test scores. *Psychometrika Monograph,* 1952, No. 7.

Lord, F. M. The relative efficiency of two tests as a function of ability level. *Psychometrika,* 1974, *39,* 351–358.

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

Pine, S. M., & Weiss, D. J. *Effects of item characteristics on test fairness* (Research Report 76–5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1976.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Rasch, G. *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielson & Lydiche, 1960.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* 1969, No. 17.

Samejima, F. A general model for free response data. *Psychometrika Monograph,* 1972, No. 18.

Samejima, F. Homogeneous case of the continuous response model. *Psychometrika,* 1973, *38,* 203–219.

Samejima, F. Normal ogive model on the continuous response level while in the multidimensional latent space. *Psychometrika,* 1974, *39,* 111–121.(*a*)

Samejima, F. *Normal ogive model on the graded response level in the multidimensional latent space.* Paper presented at the Spring Meeting of the Psychometric Society, 1974. (*b*)

Shuford, E. H., Albert, A., & Massengill, H. F. Admissible probability measurement procedures. *Psychometrika,* 1966, *31,* 125–145.

Tucker, L. R. Maximum validity of a test with equivalent items. *Psychometrika,* 1946, *11,* 1–13.

van Thillo, M., & Joreskög, K. G. *SIFASP—A general computer program for simultaneous factor analysis in several populations* (Research Bulletin 70–62). Princeton, NJ: Educational Testing Service, 1970.

Werts, C. F., Linn, R. L., & Joreskög, K. G. Intraclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement,* 1974, *34,* 25–33.

Acknowledgements

This paper, based on a doctoral dissertation completed at the University of Minnesota, was written while the author was a post-doctoral fellow in Evaluation Research at Northwestern University supported by Contract No. C-74-0115 from the National Institute of Education.

Computer time provided by the Computer Center at the University of Minnesota is gratefully acknowledged.

The author would like to express his deep appreciation to Dr. David J. Weiss, under whose supervision the dissertation on which this paper is based was completed, and to Fumiko Samejima for offering valuable suggestions in the planning of the study. Any faults that may remain are solely my responsibility.

Author's Address

Isaac I. Bejar, Psychometric Methods Program, Department of Psychology, N660 Elliott Hall, 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/