DOCUMENT RESUME

ED 253 569 TM 850 109

AUTHOR Marsh, Herbert W.; Richards, Garry E.
TITLE The Multidimensionality of the Rotter I-E Scale: An Application of Confirmatory Factor Analysis.
PUB DATE 10 Nov 84
NOTE 33p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Factor Analysis; *Factor Structure; Goodness of Fit; *Locus of Control; Males; Mathematical Models; Research Methodology; *Self Concept Measures; *Test Reliability; Young Adults
IDENTIFIERS *Rotter Internal External Locus of Control Scale

ABSTRACT

The purposes of the present investigation were to examine the dimensionality of the Rotter Internal-External (I-E) scale, to test the invariance of its factor structure before and after an intervention designed to alter I-E, and to demonstrate the use of confirmatory factor analysis. An a priori model based on the content of the Rotter items hypothesized six factors, and a review of 20 earlier factor analyses demonstrated that each of these factors had been previously identified. Results from time one and time two both indicated that this hypothesized model was able to adequately explain responses to the Rotter items, that the hypothesized factors were defined and distinguishable, that the factor structure was invariant over time, and that alternative models with fewer hypothesized factors or a simpler structure did not fit the data as well. In a more general discussion of the Rotter scale it was concluded that, when judged by current standards, the Rotter I-E scale is a poor measurement instrument: its internal consistency reliability is minimal, the rationale for its forced-choice format is dubious, the assumption of its unidimensionality is clearly wrong, and the distinguishable facets of the scale cannot be inferred with sufficient reliability to be practically useful. (Author)

ED253569

The Multidimensionality of the Rotter I-E Scale: An Application of Confirmatory Factor Analysis

Herbert W. Marsh

Department of Education, The University of Sydney

Garry E. Richards

Australian Outward Bound School

10 November, 1984

Running Head: The Rotter I-E Scale

Acknowledgments

The authors would like to the acknowledge the assistance of Jennifer Barnes, Raymond Debus, and Margaret Myers for their assistance in determining an a priori model of the factor structure underlying responses to the Rotter I-E items, and for helpful suggestions on earlier drafts of this study.

The Multidimensionality of the Rotter I-E Scale: An Application of Confirmatory Factor Analysis

ABSTRACT

The purposes of the present investigation were to examine the dimensionality of the Rotter Internal-External (I-E) scale, to test the invariance of its factor structure before and after an intervention designed to alter I-E, and to demonstrate the use of confirmatory factor analysis (CFA). An a priori model based on the content of the Rotter items hypothesized six factors, and a review of 20 earlier factor analyses demonstrated that each of these factors had been previously identified. Results from time 1 and from time 2 both indicated that this hypothesized model was able to adequately explain responses to the Rotter items, that the hypothesized factors were defined and distinguishable, that the factor structure was invariant over time, and that alternative models with fewer hypothesized factors or a simpler structure did not fit the data as well. In a more general discussion of the Rotter scale it was concluded that when judged by current standards, the Rotter I-E scale is a poor measurement instrument: its internal consistency reliability is minimal, the rationale for its forced-choice format is dubious, the assumption of its unidimensionality is clearly wrong, and the distinguishable facets of the scale cannot be inferred with sufficient reliability to be practically useful.

The Multidimensionality of the Rotter I-E Scale: An Application of Confirmatory Factor Analysis

The Internal-External Construct As Inferred From the Rotter Scale.

Internal-external (I-E) locus of control is hypothesized to be a bipolar construct; the locus is internal if a person perceives events to be contingent upon his/her behavior or relatively enduring personal characteristics, the locus is external when events are seen to be contingent upon luck, fate, the control of powerful others, the environment, or some characteristic not under his/her own control (Lefcourt, 1976; Rotter, 1966; 1975; Stipek & Wiesz, 1981). While a large number of I-E scales have been developed, the most widely used is the Rotter Scale and this instrument will be the focus of the present investigation.

The Rotter I-E scale consists of 23 pairs of statements, using a forced-choice format, and six filler questions. Each pair contains one internal statement and one external statement, and subjects make a dichotomous choice between the two alternatives. The scale is based on the assumption that I-E is a relatively unidimensional, bipolar construct. The bipolarity of responses to the Rotter scale is a necessary condition of the forced-choice format, though research with other scales suggest that the construct may not be bipolar when independent ratings are made of internal and external items (e.g., Marsh, Cairns, Relich, Barnes & Debus, 1984; also see Collins, 1974; Klockars & Varnum, 1975; Zuckerman & Gerbasi, 1977). Marsh, Cairns, et al. (1984; also see Stipek & Weisz, 1981) also demonstrated that responses may be specific to particular causes (e.g., ability, effort) and to particular situations (e.g., performance in mathematics). Nevertheless, Rotter (1966; 1975) interpretted initial research to indicate that responses to his scale were unidimensional, or at least that one general factor explained most of the variance in the total score, and this is how the scale has typically been used in I-E research.

The Present Investigation.

The primary purpose of the present investigation is to examine the dimensionality of the Rotter Internal-External (I-E) scale. This will be accomplished by reviewing previously published studies, and by reanlayzing data from Marsh, Richards and Barnes (1984). Previous research has used exploratory factor analyses (EFA), but the limitations in this approach to factor analysis render it ill-suited for this purpose. Hence, a second purpose of the present investigation is to demonstrate the application of recent advances in the use of confirmatory factor analysis (CFA) for testing hypothesized factor structures and for testing the invariance of factor structures over different sets of responses in a reanalysis of responses to the Rotter scale.

Data to be reanalyzed in the present investigation come from a study described in more detail by Marsh, Richards and Barnes (in press). Participants in that study (N=361, median age = 21, 75% male, 96% single, 60% full-time employed, 33% full-time students) completed the Rotter scale before and after the completion of the 26-day residential program called Outward Bound. Though not a major focus of the study, the authors examined responses to the Rotter scale before and after completion of the program; coefficient alpha estimates of reliability were .71 and .78, and the test-retest correlation was .68. The Outward Bound course is designed to produce a more internal orientation, and posttest scores were significantly more internal than were pretest scores ($p < .001$). For purposes of the present investigation the responses from 349 participants who completed the Rotter scale at the start (time 1) and at the end (time 2) of the 26-day program are considered.

The design and the interpretation of findings from the original study are not the focus of the present investigation, but two assumptions underlying the analysis of the Rotter scores will be investigated further. First, the use of a single total score was based on the assumption that responses to the scale are relatively unidimensional, and this assumption seems dubious. Second, the comparison of pretest and posttest scores -- for a single dimension or for multiple facets -- assumes that the factor structure was relatively invariant across the two administrations of the scale. If this invariance cannot be demonstrated, then the constructs being measured at the pretest and posttest differ, and scores representing these constructs cannot be meaningfully compared. This assumption may be particularly problematic in a study where the intervention is specifically designed to alter the I-E orientation, and may cause some facets of the construct to become more salient. It should be noted that a similar assumption of factorial invariance also underlies the meaningful comparison of responses by randomly assigned experimental and no treatment control subjects, or the comparison of responses by distinct subgroups such as males and females, blacks and whites, etc.

The Dimensionality of the Rotter I-E Scale.

The Number of Factors. A critical initial step in all factor analytic studies is to determine the number of factors needed to explain the responses. In preliminary analyses conducted for the present investigation, empirical approaches were used to determine the number of factors needed to describe responses at time 1 and at time 2. While a detailed examination of these approaches is beyond the scope of this study (see Crawford, 1975; Everett, 1983; Hakstain, Rogers & Cattell, 1982; Kim & Mueller, 1978; Marsh & Hocevar, 1984a), three approaches are most frequently used: the number of eigenvalues greater than 1.0 (Kaiser, 1960); Cattell's (1966) scree test; and a chi-square test of statistical significance (the one used here is based on Joreskog's approach to maximum likelihood factor analysis as performed by the commercially available SPSS program; see Burns, 1976). The results from each of these approaches for responses from time 1 and time 2 (Table 1) all indicate that the Rotter scale is clearly multidimensional, and that between 4 and 9 factors are needed to explain responses to the scale.

Insert Table 1 About Here

The results from previously published studies where at least one set of responses to the original Rotter scale has been factor analyzed are summarized in Table 2. The authors of each of these studies argue for the multidimensionality of the Rotter scale. In apparent contradiction to the findings from the present investigation, most studies report only two or three factors, though some report more. However, the studies that report only two or three factors typically estimate the number of factors that are needed on the basis of either an incorrect application of the eigenvalue rule that substantially underestimates the number of factors that are needed (see footnote 1), or on the interpretability of the identified factors. While the interpretability of factors is an important issue, it may not be an adequate basis for determining the dimensionality of a scale. The fact that both factors in a two-factor solution can be interpretted does not mean that there are no additional factors. Even when additional factors cannot be readily interpretted, it does not mean that a smaller number of factors can adequately explain the data. The issue of the number of factors is an important methodological issue that has not been given sufficient consideration in this research.

Though the studies in Table 2 typically did not report any of the empirical procedures listed in table 1, they often did present the proportion of variance that was explained by the factors that were identified (see Table 2). In 19 EFAs of responses to the original Rotter items, between 8% and 20% (median = 13% vs. 14% & 18% in the present investigation) of the total variance is explained by the first factor (see footnote 1), between 5% and 10% (median = 8% vs. 7% & 9% in the present investigation) of the total variance is explained by a second factor, and between 5% and 8% (median = 7% vs. 6% & 6% in the present investigation) of the total variance is explained by a third factor. In no study did two factors explain more than 30% of the variance, nor did three factors explain more than 35% of the variance. Furthermore, in all studies that reported the eigenvalues, or where they could be inferred, there were at least five greater than 1.0. These findings indicate that there is good agreement between the results of the present investigation and previous research in terms of the proportion of variance that can be explained by one, two and three factors, and that part of the apparent contradiction in the number of factors is an artifact of the manner in which the number of factors was estimated in previous research.

Insert Table 2 About Here

The Content of Factors Identified in EFA Studies. EFAs of the Rotter items summarized in Table 2 typically identified two or more of the following factors: general luck (often called general control, control ideology, or fate), political control, success via personal initiative (often called personal control), interpersonal control, control in an academic situation, and control in an occupational setting (sometimes called success mobility). Mirels (1970) described a two-factor solution consisting of a general control (called general luck here) and a political control factor, and factors like these are reported in most of the EFAs in Table 2). Gurin, Gurin, Loa and Beattie (1969), though not based on the original set of Rotter items, reported a personal control factor in addition to the luck and political control factors. Sanger and Alker (1972) described this factor as tantamount to endorsing the view that one's own efforts are responsible for one's success or failure, and it is called success via personal initiative in the present investigation. While such a factor has apparently been found in several studies that consider only the original set of Rotter items, its interpretation is sometimes tenuous. The factors characterized here as general luck and success via personal initiative are sometimes difficult to distinguish, and may have a considerable overlap in the items that define them. Furthermore, when the two are not identified as separate factors, they are typically incorported into one factor that is usually called general control that appears to have components of each. This situation is exacerbated by the fact that many Rotter items ask respondents to choose between statements that refer to luck and success due to personal initiative.

Abrahamson, Shulderman and Shulderman, (1973) reported a three-factor solution, but their third factor was an interpersonal or social factor that has been identified in a number of other EFAs. Particularly when more than three factors are considered, some researchers have found an academic control factor (e.g., Garza & Widlak, 1977; Little, 1977; Watson, 1981; and perhaps O'Brien & Kabanoff, 1981). Finally, an occupational factor has been identified in some EFAs (O'Brien & Kabanoff 1981; also see Gurin, Gurin & Morrison, 1978), though the factor is defined by only two items and thus is necessarily weak. While most researchers attempted to interpret only two or three factors, four studies have identified either five (Garza & Widlak, 1977; Little, 1977; Watson, 1981) or six (O'Brien & Kabanoff, 1981) factors that are apparently similar to those described here.

Despite the apparent consistency of the findings summarized in Table 2, the conclusion drawn by Dixon, McKee and McRae (1976, p. 318) is still valid: even though researchers often use similar descriptive labels for their factors, the actual item content of their factors may vary dramatically from study to study. This is particularly true in these EFAs in that: most loadings are small, some items do not load substantially on any factor, some items load on more than one factor, and the size of a loading used to decide whether or not an item loads significantly on a particular factor is typically arbitrary.

In reviews of this factor analytic research, Lange and Tiggemann (1981; also see Cherlin & Bourque, 1974) suggest that the use of the Rotter I-E items in the context of additional items, and its administration to highly selected samples may influence the instrument's factor structure. Many researchers have attempted to compare factor analytic results obtained from different sets of responses, or to compare their factor solution with those obtained in other studies. Comparisons of different factor structures within a single study are most frequently made for responses by males and by females, by a general population and by some more narrowly defined subgroup, or by the same group of subjects on more than one occasion. Methodological limitations inherent in the use of EFA for purposes of testing factorial invariance in such comparisons are discussed latter, and render the results of such exercises as problematic. Nevertheless, most researchers suggest that at least some, and in some cases, all, of their factors generalize across different analyses.

Other researchers have factor analyzed responses to the scales where the 23 forced-choice Rotter items are presented as 46 Likert-type items, thus eliminating some of the problems inherent with the forced-choice format (e.g., Collins, 1974; Klockars & Varnum, 1975; Zuckerman & Gerbasi, 1977). While such an approach is reasonable, and these results appear to be consistent with the general findings described here, a detailed examination of these studies is beyond the scope of the present investigation.

Since studies summarized in Table 2 include only those that used the original Rotter items, the important study by Gurin, et al., (1978) was not included. They eliminated the three Rotter items that refer to academic situations and included three additional items in a study based on a large representative sample of the general adult population. In each of the separate analyses of responses by white men, by white women, and by blacks they found seven eigenvalues greater than 1.0, but chose to rotate five factors. While there was not complete agreement in the factors identified in their three subpopulations, they report five factors (control ideology, political control, personal control, interpersonal control, and success mobility) that are similar to five of the six (general luck, political control, success via personal initiative, interpersonal control, and occupational situations) summarized earlier; they did not find an academic factor since they had excluded the items that would have identified this factor. While they were able to clearly identify these distinguishable factors, the coefficient alphas for their scales were consistently low (.41 to .65, mean = .55), due at least in part to the short length of some of the scales. They then went on to relate different facets of the the I-E construct to a variety of other constructs in an attempt to explore their convergent and discriminant validity. Thus, while the use of a somewhat different set of items may call into question of generality of this study to other research summarized here, it represents an important contribution to the study of the dimensionality of the I-E construct.

The Application of Confirmatory Factor Analysis (CFA).

In each of the studies described earlier researchers have used exploratory factor analyses (EFA) to examine the factor structure of responses to the Rotter I-E scale for a single group, or to compare factor structures derived from different sets of responses. While such an approach may be heuristic and suggest plausible factors, it is generally unacceptable. With EFA the researcher has little control over the factor structure to be tested beyond the number of factors to be rotated and perhaps the degree of correlation among the factors. Hence, the researcher is unable to test the ability of a hypothesized structure to fit the data and is limited to an a posteriori interpretation of the factor structure that is derived by the EFA procedure. Since the EFA solution is not identified, there are an infinite number of alternative solutions that are mathematically equivalent (i.e., that fit the equally well and differ only in terms of the rotation of the dimensions) and some may result in quite different interpretations of the underlying factors. EFA provides no indication of how well a hypothesized factor structure would be able to fit the data, nor how alternative structures compare in their ability to fit the data. Finally, the inability to define the model to be tested for any one set of responses makes the comparison of solutions derived from different sets of responses so problematic that Alwin and Jackson argue that "the use of exploratory factor analysis in its conventional form to examine issues of factorial invariance is of limited utility " (1981, p. 253; also see Marsh & Hocevar, 1984a). In contrast to EFA, CFA allows the researcher to: define and test a hypothesized factor structure; uniquely estimate the parameters used to define a hypothesized model; examine a model's ability to fit the data; compare the goodness-of-fit for alternative models; and test the invariance of all parameters, or any subset of parameters, used to define a model across different sets of responses. (For a more detailed comparison of EFA and CFA, and a general introduction to CFA, see Bagozzi, 1980; Huba & Bentler, 1982; Marsh & Hocevar, 1983; 1984a; 1984b; in press; Joreskog, 1971; 1980; Joreskog & Sorbom, 1981; Long, 1983a; 1983b; and Pedhauzur, 1982).

The Creation of an A Priori Model. CFA, unlike EFA, begins with a specifically defined model. When the factors that an instrument is designed to measure are well-defined, then the design of the instrument serves as one a priori basis for defining the factor structure. For example, a single-factor model where all items are allowed to load on one factor is consistent with the design of the Rotter scale. However, the formulation of models when there is no clearly defined a priori model, or when alternative models are sought, is not straight-forward. One possible approach is to first use EFA to suggest viable structures and then to test these structures with CFA. Alternatively, past research and the content of the items can be used to formulate alternative models.

In the present investigation, an inspection of the 23 Rotter items by the first author suggested six possible factors: General Luck, Political Control, Success Via Personal Initiative, Interpersonal Control, Academic Situations, and Occupational Situations. The first four represent different sources of control, each appeared to be represented by at least five items, and each appeared to be similar to factors identified in previous research. The last two factors are defined by specific situations rather than sources of control; they are apparently represented by only 3 and 2 items respectively, and they also have been identified in previous research. In order to generate an a priori model, the first author and three colleagues independently classified each item into one or more of these six categories. (Since each Rotter item consists of two separate statements, many items cannot be unambiguously to categorized into just one category and when an item was classified into more than one category, the category that best represented the item was also indicated.) Agreement among the four sets of classifications varied from 85% to 95% (mn = 90%). Based on these responses, a single "best" category and secondary categories (any category indicated by at least two of the four raters) were designated for each item. The results of this process provided good support for five, and some support for the sixth, of the six proposed categories; the Occupational Situation factor was defined by only two two items and neither of these items were perceived as best represented by this category by any of the raters. The results of this exercise were used to specify a series of a priori models that hypothesized 1 to 5 factors to explain responses to the Rotter items (see Appendix I).

The Definition of CFA Models. In CFA performed by LISREL, alternative models are specified in terms of three design matrices that are conceptually similar to matrices resulting from EFA: LAMBDA Y, a matrix of factor loadings; PSI, a factor variance/covariance matrix that represents relations among the factors; and THETA EPSILON, a matrix that contains error/uniquenesses -- similar to one minus the communality estimates in EFA -- in the diagonal. In the definition of each model, every element in these matrices is: a) fixed at a predetermined value, usually 0 or 1,; b) freely estimated as part of the analysis; or, c) constrained such that two elements have the same value (this type of constraint is used to test for the invariance of estimates for two different sets of responses as described below). For purposes of the present investigation, alternative models proposed 1, 2, 3, 4, or 5 factors corresponding to the categories described earlier (see Appendix I). For all but the one-factor model, simple and complex structures were examined. For simple structures each item was allowed to load on only one factor, while for complex structures some items were allowed to load on more than one factor as specified in the a priori models. In order to define each model, one item was selected to serve as a reference indicator for each factor; its factor loading was fixed to be 1.0 on that factor and it was not allowed to load on any other factor (see Long, 1983a, pp. 49-55 for further discussion of reference loadings).

Goodness-of-fit. A plethora of indices of a model's goodness-of-fit are used in CFA, but there are no well established guidelines for what minimal conditions constitute an adequate goodness-of-fit. In general: a) the parameter estimates are examined in relation to the substantive model and permissible values; b) the chi-square value for a model is evaluated in terms of statistical significance and compared with that obtained for alternative models; and, c) subjective indices of goodness-of-fit are examined and compared with values obtained for alternative models. In contrast to traditional significance testing, the researcher may prefer a nonsignificant chi-square test of statistical significance that indicates that a proposed model is able to fit the observed data within the limits of chance. However, there are problems with this test. First, the test is highly sensitive to departures from multivariate normality. Second, in large complex problems where the sample sizes and the number of measured variables are large, the test is so powerful that the chi-square will generally be statistically significant even when a model provides a reasonably good fit to the data. Hence, most practical applications of CFA require a subjective evaluation of whether a statistically significant chi-square is small enough to constitutes an adequate fit.

Many subjective indices of goodness-of-fit have been developed and some of the most frequently used have been calculated for models described in this study. A detailed examination of these indices is beyond the scope of this study (see Bentler & Bonett, 1981; Cliff, 1983; Fornell, 1983; Joreskog & Sorbom, 1981; Long, 1983a; 1983b; Marsh & Hocevar, 1984a; 1984b; in press) and discussion will be limited primarily to the chi-square/df ratio and the Tucker-Lewis Index (TLI). Researchers generally interpret X^2/df ratios of less than 2.0 to represent an adequate fit, but the value of this ratio will increase with sample size independent of the ability of the model to fit the data. Because of this problem, alternative indices such as the TLI were devised that are not dependent on sample size. The TLI scales the observed chi-square along a 0-1 scale where 0 represents the chi-square obtained from a null model and 1 represents an ideal fit. The logic of this index is similar to indices such as eta and omega that are frequently used in ANOVA as an indication of the proportion of variance explained. Since there is not wide agreement for evaluating the absolute values of these different indices, they are more useful, perhaps, as a basis for comparing alternative models used to describe the same set of responses.

Insert Table 3 About Here

The Results of Alternative Models. The results of alternative models are summarized in Table 3. In general, models that hypothesize more factors are better able to fit the data than models that propose fewer factors, and complex models are better able to fit the data than simple models. This pattern is somewhat stronger for data from time 2 than for time 1. Nevertheless, for both time 1 and time 2, the five-factor model with a complex structure (Model 5B) is better able to fit the data than models that hypothesize fewer factors or that propose a simple structure of factor loadings. The chi-square for Model 5B is statistically significant for data for time 1 and time 2, but the goodness-of-fit indices suggest that the model is able to adequately fit the data. In particular, despite the relatively large sample size, the X^2/df ratios (1.24 for time 1, and 1.35 for time 2) are much smaller than 2.0. Although the chi square value for Model 5B is slightly smaller for time 1 than time 2, the TLI is better for time 2. This occurs because the average covariation among items is larger for time 2 than time 1, and thus there is more variance to be explained in those responses.

A sixth factor, Occupational Situations, was originally proposed, but since only two items (9 and 13) were classified into this category and neither was perceived to be primarily determined by this factor, it could not be adequately defined. However, this factor is represented in Model 5C by allowing the error/uniqueness terms (in THETA EPSILON) for these two items to be correlated. Even though Model 5B provided a good fit to the data, and only one additional parameter is estimated in Model 5C, the chi-square for Model 5C is significantly better. This means that in both sets of responses the two occupational items are more highly correlated with each other than can be explained in terms of the five factors. This is interpretted as support for an Occupational factor, though an alternative interpretation would be that these two items are influenced by a method/response bias produced by the reference to an occupational setting that appears in both items.

Tests of Factorial Invariance.

Each of the models summarized in Table 2 was tested in terms of its ability to fit the data for time 1 and for time 2. Model 5C provided the best fit to the data for both sets of responses, but none of the parameter estimates were constrained to be equal across time 1 and time 2 in order to test the invariance of the hypothesized structure. For purposes of the present investigation, the invariance of Models 1 and 5C is tested. Model 5C provided the best fit to the data, and so it is relevant. However, since most research uses a single total score to summarize responses to the Rotter scale -- the procedure consistent with the original design of the instrument, tests of the invariance of the one-factor model may also be relevant.

For tests of the invariance of the factor structure over two different sets of responses, the estimated values for any single parameter, for any subset of parameters, or for all parameters can be constrained to be equal in solutions representing each set of responses. The LISREL procedure fits the data subject to these equality constrain , and an evaluation of the goodness-of-fit indices is used to infer how de ental the constraints are to the fit of the data (for a more detailed discussion of factorial invariance see Alwin & Jackson, 1981; Everett & Rog, 1981; Joreskog & Sorbom, 1981; Marsh & Hocevar, in press; 1984a). To the extent that a model with equality constraints is able to fit the data nearly as well as a model without these constraints, there is support for the invariance of the constrained parameters. It should be noted that the chi-square value for a model with equality constraints can never be smaller than that of the corresponding unconstrained model, and that the chi-square values will only be the same if the parameter estimates under consideration are exactly equal when no constraints are imposed. In this sense, the unconstrained model represents an upper limit, or target, for the goodness-of-fit for the constrained model. An alternative index of goodness-of-fit for the constrained model, the target coefficient (TC), is the ratio of the chi-square values for a constrained and an unconstrained models. The TC provides an estimate of the proportion of variance in the unconstrained model that a model with equality constraints is able to explain (see Table 4, also see Marsh & Hocevar, in press, for a similar derivation).

In the present investigation, four tests of factorial invariance were conducted for Models 1 and SC (see Table 4). In test 1, the most demanding test, every parameter estimate for time 1 was constrained to be equal to the corresponding parameter estimate for time 2 -- factor loadings in LAMBDA Y, factor variances and covariances in PSI, and error/uniquenesses in THETA EPSILON. Progressively less demanding tests of invariance were examined such that elements in THETA EPSILON (tests 2, 3 & 4), elements in PSI (tests 3 & 4), and elements in LAMBDA Y (test 4) were not constrained to be equal over time. In test 4, the least restrictive model, no invariance constraints were imposed at all; the chi-square and df for this model are the sum of the values determined separately for responses from time 1 and from time 2 (Table 3). For both Models 1 and SC, test 1 produced poorer goodness-of-fits than did the less restrictive tests of invariance, though even here the fits were reasonably good and the TCs were over .90. In tests 2 and 3 where the elements in THETA EPSILON were not constrained to be equal, the goodness-of-fit indices differed little from test 4, and the [illegible] were all .96 or greater. These findings provide strong support for the invariance of factor loadings, factor variances, and factor covariances over time. As described earlier, this finding is important in that it provides support for the comparison of pretest and posttest scores, as well as providing further evidence of the generality of the factor structure.

Insert Table 4 About Here

The interpretation of the tests of invariance for Model SC is straightforward in that: a) the constrained models are able to fit the data, and b) the goodness-of-fit indices for the constrained models, except possibly for test 1, differ little from the unconstrained model. However, the interpretation of the invariance tests for Model 1 is more complicated. As with Modal SC, the goodness-of-fit for the constrained models differ little from the unconstrained models. However, none of the tests of Model 1 provides nearly as good a fit as do the tests of Model SC. This situation is illustrated by the fact that the TCs for the tests of Model 1 are very high, even though the other goodness-of-fit indices are much poorer than those for Model SC. For purposes of the present investigation, this situation is interpreted to mean that the responses to the Rotter items cannot be adequately explained by a single score, but that if a single score is used then it is reasonable to compare it at time 1 and time 2. However, the justification for the use of a single score rests in part on additional theoretical considerations, and these will be discussed latter.

<u>An Examination of the Parameter Estimates.</u>

In order to evaluate Model SC it is important to examine the values of he actual parameter estimates (see Table 5). These estimates represent responses from both time 1 and time 2 in that the estimates come from the test of invariance where all the estimated parameters were required to be equal over time. A preliminary inspection reveals that no parameter estimates fall outside of their permissible limits -- factor variances or error/uniquenesses that are negative, or factor covariances that are larger than the product of the factor standard deviations (i.e., factor correlations greater than 1.0). By itself this observation provides only minimal support for the model, but such violations are frequent in CFAs with complex factors and argue against the validity of the proposed model. The parameter estimates do not resemble those from EFA studies, partly because the estimates were based on the item covariances instead of the item correlations, and partly because reference variables were used to define each factor. However, for each estimated parameter -- those not fixed at a predetermined value -- the LISREL program also provides a standard error of the estimate that is helpful in the interpretation of the results. This standard error can be used to form a t-ratio and to infer statistical significance; if the ratio is more than 1.96 the estimate differs significantly from zero.

Insert Table 5 About Here

The t-ratios for all five factor variances (the diagonal of PSI) are greater than 3, indicating that each of the factors describes a significant portion of variance. An examination of the factor loadings (in LAMBDA Y) indicates that each item contributes significantly to at least one of the factors that it was hypothesized to represent. In some cases where an item was hypothesized to define more than one factor not all the loadings are statistically significant, but the first four factors are each represented by at least four items that are significant as well as by their reference variable. The fifth factor, Academic Situations, is only defined by three items -- a reference item and two other items -- but the t-ratios for all the estimated factor loadings and the factor variance are statistically significant. Hence, even the Academic Situation factor is clearly identified though it would be preferable to have more items. The third factor, Success Via Personal Initiative, is defined by 13 items, but estimates for seven of these failed to reach statistical significance. (This was also he factor that produced the least agreement among raters in the preliminary exercise used to formulate the a priori model, and in earlier EFAs described in Table 2.) An examination of the content of the items hypothesized to define the third factor did not reveal any obvious differences between items with and without significant loadings (e.g., those written in the first and third person). However, except for the reference variable, every item hypothesized to define this factor was also hypothesized to reflect one or more other factors in the a priori model, and had a statistically significant loading on at least one other factor in the empirical findings. Hence, the difficulty in determining what items best define this factor may be due to the design of the Rotter instrument.

The factor covariances in PSI are all statistically significant, indicating that the factors are correlated. When these covariances are transformed into correlations by dividing each by the product of the factor standard deviations that they represent, the correlations (see Table 5) vary from .2 to .8. Only the correlation between General Luck and Success Via Personal Initiative is greater than .55, and these are the two factors with the greatest item overlap in the a priori model. The fact that many items ask respondents to select between one statement referring to luck and one referring to success via personal initiative means that it is difficult to distinguish between these factors on the basis of the Rotter items.

Joreskog and Sorbom (1981) describe the modification index (MI) that is useful in the further evaluation of a model. For every parameter that is fixed or constrained, the MI provides a lower-bound decrease that would result in the chi-square value if that parameter were freely estimated. Joreskog and Sorbom (1981) suggest that the MI for a parameter should be at least 5 before a researcher considers modifying the hypothesized model; a change in the chi-square of 5 is statistically significant at $p < .01$, though the probablity of a type I error is grossly inflated when such a test is applied to a large number of parameters (also see Marsh & Hocevar, in press; 1984). In the present investigation, 72 of the 115 factor loadings for Model 5C were hypothesized to be zero indicating that a particular item would not load on a particular factor, and the validity of these hypotheses can be tested with the MI for each of these parameters. In various tests of the invariance of the parameter estimates for Model 5C, up to 80 parameters were constrained to be equal, and each MI indicates how much the contraint of the one parameter affected the overall chi-square. For Model 5C, MIs were less than 5 for all factor loadings, factor variances and factor covariances in all tests of invariance for data from time 1 and time 2. In test 1 of the factorial invariance of Model 5C where elements in THETA EPSILON were constrained to be equal over time, approximately 15% of the error/uniqueness terms -- though not the one correlated error term used to represent the occupational items -- were larger than 5. These results provide further support of the interpretation of invariance tests presented earlier, and also suggest that no item would have contributed significantly to a factor that it was not hypothesized to represent.

Summary and Implications

The primary purpose of the present investigation was to determine the dimensionality of the Rotter I-E scale by reviewing previous research and the methodology of that research, and by applying CFA procedures that incorporate recent advances in the application of factor analysis. Though the scale was originally designed to be unidimensional and it is typically summarized with a single score, nearly all studies that have specifically examined its dimensionality have concluded that it is multidimensional. Empirical guidelines to the number of factors that are needed, the results of previous research, an a priori model based on the content of the Rotter items, and the results of the CFA suggest that the Rotter scale contains 5 or 6 distinguishable factors. Though many earlier studies report only 2 or 3 factors, an examination of the basis for the decision to report so few factors suggests that additional factors were needed to adequately explain the responses to the Rotter items.

CFA, unlike EFA, begins with an a priori model of the hypothesized factor structure. An a priori model based on the content of the Rotter items, Model 5C, hypothesized five distinguishable factors and incorporated a sixth factor through the use of correlated error/uniquenesses. CFA results from time 1 and from time 2 both indicated that this five-factor model was able to explain adequately the responses to the Rotter items, and that hypothesized models with fewer factors or a simpler structure did not fit the data as well. An examination of the parameter estimates generated by Model 5C indicated that each of the factors was reasonably well defined, was distinguishable from the other factors, and explained a significant portion of the variance. Every item contributed significantly to at least one of the factors that it was hypothesized to represent, and no item contributed significantly to a factor that it was not hypothesized to represent. Tests of the invariance of the factor structure indicated that the factor loadings, factor variances, and factor covariances, though perhaps not error/uniquenesses, were invariant over the two administrations of the test.

Comparisons among factor structures identified in previous research must be made cautiously because of the limitations inherent in the use of EFAs employed in those studies. In no other study known to the authors was CFA used to test any a priori model of the factor structure of the Rotter items. Rather, support for a particular model was based on an idiosyncratic, a posteriori examination of factor loadings derived from an EFA. A test of factorial invariance with CFA is the recommended statistical procedure, perhaps the only justifiable procedure, for comparing the factor structures from two different sets of responses based on: a) before-after scores for a single group of subjects; b) responses by randomly assigned treatment and no treatment subjects; c) responses by different subgroups within a larger population; and d) responses to an original set of items and responses to the same set of items embedded in the context of additional items. Hence, this procedure provides a basis for examining many questions raised in this area of research. More importantly the comparison of mean scores in each of these comparisons, the basis of nearly all experimental and correlational research, rests on the rarely tested assumption that the factor structure for the different sets of responses is relatively invariant. Consequently, the procedures described in the present investigation represent an important methodological advance over previous research -- one that can be used in reanalyses of data from earlier studies to examine previously unanswered questions and untested assumptions. The apparent success of the a priori model also provides support for the classification of of items into categories by independent raters as one basis for deriving a priori models in other other CFA studies where well-defined factor structures do not exist.

Rotter's intention was to design a general I-E measure, though he recognized that there may be many distinguishable facets and situationally specific components of the construct. In order to accomplish this purpose Rotter constructed items that reflected many different specific facets with a relatively small number of items, and the surprisingly small average correlation among the responses to the Rotter items -- about .1 -- suggests that he was successful. While it may be justifiable to use such an approach to infer a general component, one that is superordinate to the specific components on which it is based, such an approach will not produce a unidimensional scale. Well-defined unidimensional scales are typically based on a set of relatively homogeneous items that result in responses that are substantially correlated -- exactly opposite to the approach used by Rotter. Rotter's approach also complicates the identification of a clearly defined factor structure. Even though the total score may reflect a broad range of specific components, many of these components may be represented by a small number of items or even a single item. Ironically, this difficulty has led some researchers to conclude that only one, or at most a few, facets are being measured by the Rotter scale. The fact that only a few factors are sufficiently well defined by the Rotter items to be consistently identified in EFA studies does not mean that a few factors are able to explain responses to the items. Empirical approaches used to estimate the number of factors represented by the Rotter items, as well as the logic used in the construction of the scale, suggest that there are many facets. Indeed, it is the extreme heterogeneity of the Rotter items rather than their homogeneity that makes it difficult to identify the many facets that are apparently reflected in the total score.

Part of the confusion in research based on the Rotter scale stems from an apparent ambiguity in how a general scale is defined, and this problem is not specific to locus of control research. A similar predicament in self-concept research has been recent examined by Marsh (Marsh, 1984; Marsh & Shavelson, in press). Historical , self-concept was typically inferred from responses to a hodge-podge of self-referent items, and researchers' inability to adequately identify distinguishable facets of self-concept was taken to mean that self-concept was a unidimensional construct representing a general self-concept. However, subsequent theoretical models (e.g., Shavelson, Hubner & Stanton, 1976), and factor analyses of responses to instruments derived from these models (e.g., Marsh & Shavelson, in press; Marsh, Barnes & Hocevar, in press), now provide strong support for the multidimensionality of self-concept. In this approach, facets were hypothesized on the basis of a well-articulated theoretical model, a sufficient number of items were constructed to adequately define each separate facet, and then empirical tests were conducted to establish that the hypothesized factor structure was able to fit the responses and to examine other psychometric properties. The advantages of such an approach are that the underlying facets are specifically hypothesized and are empirically testable, thus providing a clear basis for interpretting responses to the instrument. Also, external criteria are more accurately predicted by specific facets that are logically related to them than by a general total score. This approach has been used with apparent success in self-concept research, and it may also be useful in locus of control research (see Gurin et al., 1978).

Even if the specific facets underlying a general I-E construct, and their structure, can be articulated and supported empirically, this may not resolve the issue of how a general I-E score should be inferred. Marsh (1984), again based on self-concept research, described three different approaches: a) the total score from an idiosyncratic set of specific items that may or may not be balanced with respect to the appropriate facets of the general construct; b) a higher-order factor, one derived from a higher-order factor analysis of relations among well-defined first-order facets that is consistent with a hierarchical model of the general construct; or c) a superordinate, relatively unidimensional component of the general construct that is inferred from responses to items that are not specific any particular facet (in self-concept research this is sometimes referred to as self-esteem). Rotter's strategy appears to correspond to the first approach, though the second approach would seem to be psychometrically more defensible and perhaps more heuristic. However, some researchers (e.g., Cherlin & Bourque, 1974; Lange & Tiggemann, 1981) suggest that the general control facet identified in many EFAs may represent a general factor like that suggested in the third approach, and this possibility deserves further consideration.

In locus of control research, as in other areas of research, it may be ill-advised to infer a general construct from a small number of items designed to represent many specific components. First, the particular items are rarely based on a theoretical model that provides a basis for sampling specific components. Without such a model, there is no basis for evaluating whether the items reflect the general theoretical construct, or whether they are balanced in respect to the appropriate facets of the general construct. Second, even if the selection of items is based on a well-defined model of the general construct, it is difficult to determine empirically whether these facets are reflected in responses to the relatively small number of items chosen to represent these facets. A better approach is to construct instruments to measure each of the facets derived from a theoretical model, and to use empirical procedures such as CFA to test if these facets are reflected in responses to the instruments. Once a well-defined set of lower order factors has been identified, hierarchical CFA can be used infer more general facets from the lower order facets (see Marsh, 1984; Marsh & Hocevar, in press). On the basis of such research it may then be possible to select a relatively small number of items that adequately reflect the general construct, though this possibility would require further research.

What are the implications of this research for the status of the Rotter scale and future research? The present investigation clearly shows that the Rotter scale is multidimensional, as does nearly all research that has examined this issue, and so there is no justification for the use of the Rotter scale as a unidimensional construct. However, even though multiple facets are clearly identified, most are not sufficiently reliable to be considered separately -- their practical application would require the construction of new scales that contain more items and items that are more clearly related to the specific facet that each is designed to measure. Perhaps, more in keeping with the original design of the Rotter scale, its total score might be used as a higher-order general I-E measure that is based on an average of specific facets of the construct. However, the justification for this must reside in a theoretical model that specifies what the particular facets are, how they combine to form the general construct, and how these characteristics are reflected in the sampling of items in the Rotter IE scale. Clearly, such empirical support does not exist for the Rotter scale, and it seems unlikely that the particular mix of items that appears on any instrument would adequately satisfy these criteria unless such a strategy were used in the original construction of items.

Other potential problems for the use of the Rotter scale stem from the forced-choice format of the Rotter items. First, though not the focus of the present investigation, the assumption that internality and externality represent bipolar ends of a single continuum -- that the correlation between these two constructs approaches -1.0 -- is problematic. Second, the fact that each item is comprised of two statements that typically refer to different facets of I-E means that the determination of its factor structure is unduely complicated (see Collins, 1974; Klockars & Varnum, 1975; Zuckerman & Gerbasi, 1977). Only items from the Political Control facet are generally comprised of two statements from the same facet, and this may explain why this facet is nearly always identified as a separate facet. Finally, once the multidimensionality of the construct is recognized, a host of new theoretical issues about how statements are paired to form each item must be examined. Logically, it would seem that both the internal and external statements that comprise each item should represent the same facet, though other strategies may be viable. The authors' personal recommendation is against the use of a forced-choice format in locus of control research. However, even if other researchers favor the use of a forced-choice format, this most be justified on the basis of further theoretical and empirical development of the I-E construct.

The research conducted by Rotter more than 20 years ago, as summarized in his 1966 monograph, and the I-E scale that was derived from that research, represented an important basis for locus of control research. The scale has been very heuristic in terms of the amount of research that it has stimulated. Nevertheless, judged by current standards, the Rotter I-E scale is a poor measurement instrument: the internal consistency reliability of its total score is minimal, the rationale for its forced-choice format is dubious, the assumption of its unidimensionality is clearly wrong, the distinguishable facets of the scale cannot be inferred with sufficient reliability to be practically useful, and there is no theoretical basis for combining responses to the different facets to form a general I-E score. The continued reliance of locus of control research on this type of instrument is likely to impede further development of the construct, particularly its multifaceted nature. Instead, researchers should develop systematic models of the important facets of the general I-E construct and, perhaps, a hierarchical ordering of these facets. Such a model should then be the basis for the construction of appropriate measurement instruments. In this way, empirical results can be used to support, refute, or revise both the instrument and the theory. The original research summarized by Rotter and subsequent research using the Rotter I-E scale provide one important basis for further research and instrument construction, but the continued use of the Rotter I-E scale is not recommended.

FOOTNOTES

1 -- The eigenvalue rule is based on the number of principal components that have eigenvalues greater than 1.0 -- the number of unrotated factors that are able to explain at least 4.3% (i.e., 1/23 x 100% where 23 is the number of Rotter items) of the total variance in analyses of the Rotter items. Some early studies in this area, apparently in a mistaken interpretation of the eigenvalue rule, have applied the rule to eigenvalues resulting from the rotated factors and the common factor variance, and this approach has then been used in some subsequent replications (see Gurin, et al., 1978, for further discussion). There is no justification for this second approach offered in any of the studies where it is used, suggesting that it does represent a misunderstanding of of the original rule, and it results in a substantially smaller number of eigenvalues that are greater than 1.0 than the correct application of the eigenvalue rule. The apparent confusion between total variance and common variance has also been a source of confusion in other aspects of this research as described in footnote 2.

2 -- Rotter (1966; 1975) reported that in two factor analyses, conducted early in the development of his scale, most of the variance could be accounted for by a single factor; he reported a figure of 53% for one of the analyses but did not give details about how this figure was derived. However, several sources of evidence indicate that Rotter must not have been referring to the proportion of the total variance that could be explained by a single factor. First, his suggestion is completely out of the range of findings summarized in Table 2 -- in 19 factor analyses a single factor factor explained no more than 20% of the variance (median = 13%) and not even three factors were able to account for more than 35% of the variance. Second, the amount of variance that can be explained by a single factor is closely related to the average correlation among items or, equivalently, the coefficient alpha estimate of reliability for the total score. However, the average correlations among items in the two studies cited by Rotter, derived from the internal consistency estimates reported in his 1966 monograph, were .092 and .088, and are <u>lower</u> than values in the present investigation (.096 for time 1 and .135 for time 2). This finding argues that less than 14% of the variance in the two studies cited by Rotter could be explained by a single factor, and leaves in doubt the derivation of the figure actually reported by Rotter. Finally, in a reanalysis of the original correlation matrix from one of the original studies cited by Rotter, Prociuk (1977) found that the no more than 9% of the variance could be explained by the first factor rather than the 53% reported by Rotter. Prociuk suggested that the 53% represented the percentage of common variance and not the total variance that could be explained by a general factor.

REFERENCES

Abrahamson, D., Schuldermann, S. & Schuldermann, E. Replication of dimensions of locus of control. Journal of Consulting and Clinical Psychology, 41, 320.

Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. D. Jackson & E. F. Borgotta (Eds.), Factor analysis and measurement in sociological research: A multidimensional perspective. Beverly Hills, Calif.: Sage.

Bagozzi, R. P. (1980). Causal models in marketing. New York: Wiley.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Burns, P. (June, 1976). SPSS subprogram JFACTOR -- Joreskog Factor Analysis (Document No. 412). In SPSS Supplement for SPSS-6000 version 6.5. Vogelback Computing Center, Northwestern University, Evanston, IL.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Cherlin, A. & Bourque, L. A. (1974). Dimensionality and reliability of the Rotter I-E scale. Sociometry, 37, 565-582.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. Multivariate Behavioral Research, 18, 115-126.

Collins, B. (1974). Four components of the Rotter Internal-External scale: Belief in a difficult world, a just world, a predictable world, and a politically responsive world. Journal of Personality and Social Psychology, 29, 381-391.

Crawford, C. B. (1975). Determining the number of interpretable factors. Psychological Bulletin, 82, 226-237.

Dixon, P. N., McKee, C. S., & McRae, B. C. Dimensionality of three adult objective locus of control scales. Journal of Personality Assessment, 40, 310-319.

Everett, J. (1983). Factor Comparability as a means of determining the number of factors and their rotation, Multivariate Behavioral Research, 18, 197-218.

Everett, J. & Rog, D. J. (1981). The relevance of structural invariance. Paper presented at the 1981 Annual Meeting of the American Educational Research Association, Los Angeles.

Fornell, C. (1983). Issues in the application of covariance structure analysis. Journal of Consumer Research, 9, 443-448.

Garza, R. T. & Widlak, F. (1977). The validity of locus of control dimensions for Chicano populations. Journal of Personality Assessment, 41, 635-643.

Gurin, P., Gurin, G., Lao, R. C., & Beattie, M. (1969). Internal-external control in the motivational dynamics of Negro youth. Journal of Social

Issues, 25, 29-53.

Gurin, P., Gurin, G., & Morrison, B. M. (1978). Personal and ideological aspects of internal and external control. Social Psychology, 41, 275-296.

Hakstain, A. R., Rogers, W. T. & Cattell, R. B. (1982). The behavior of number-of-factor rules with simulated data. Multivariate Behavioral Research, 17, 193-219.

Huba, G. J., & Bentler, P. M. (1982). On the usefulness of latent trait modeling in testing theories of naturally occurring events. Journal of Personality and Social Psychology, 43, 604-611.

Joreskog, K. G. (1980). Structural analysis of covariance and correlation matrices. Psychometrica, 43, 443-477.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrica, 36, 409-426.

Joreskog, K. G. & Sorbom, D. (1981). LISREL V: Analysis of Linear Structural Relations By the Method of Maximum Likelihood. Chicago: International Educational Services.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 1541-151.

Kim, J. & Mueller, C. W. (1978). Factor analysis: Statistical methods and practical issues. Beverly Hills, CA: Sage.

Klockars, A.J. & Varnum, S.W. (1975). A test of the dimensionality assumptions of Rotter's Internal-External scale. Journal of Personality Assessment, 39, 397-404.

Lange, R. V., & Tiggemann, M. (1981). Dimensionality and reliability of the Rotter I-E locus of control scale. Journal of Personality Assessment, 45, 398-406.

Lefcourt, H. M. (1976). The locus of control: Current trends in theory and research. New York: Wiley.

Little, C.B. (1977). Dimensionality of the Internal-External Control scale: Some further evidence. Journal of Genetic Psychology, 131, 329-330.

Long, K. S. (1983a) Confirmatory factor analysis: A preface to LISREL. Beverly Hills, CA: Sage.

Long, K. S. (1983b) Covariance structure models: An introduction to LISREL. Beverly Hills, CA: Sage.

Marsh, H. W. (1984). The hierarchical structure of self-concept: An application of confirmatory hierarchical factor analysis. (In Review) Department of Education, University of Sydney, Australia.

Marsh, H. W., Barnes, J., & Hocevar, D. (in press). Self-other agreement on multidimensional self-concept ratings: Factor analysis and multitrait-multimethod analysis. Journal of Personality and Social Psychology, Marsh, H.W., Cairns, L., Relich, J., Barnes, J., & Debus, R.L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. Journal of Educational Psychology, 76, 3-32.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H. W., & Hocevar, D. (1984a). The comparison of factors from different groups with exploratory and confirmatory factor analysis. (In Review).

Marsh, H. W. & Hocevar, D. (1984b). The factorial invariance of students' evaluations of college teaching. American Educational Research Journal.

Marsh, H. W. & Hocevar, D. (in press). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. Psychological Bulletin.

Marsh, H. W., Richards, G., & Barnes, J. (in press). Multidimensional self-concepts: The effect of participation in an Outward program. Journal of Personality and Social Psychology.

McDonald, A. P. (1973). Measures of internal-external control. In J.P. Robinson & P.R. Shaver (Eds.), Measures of social psychological attitudes (pp. 169-243). Ann Arbor: Institute for Social Research, University of Michigan.

Mirels, H. L. (1970). Dimensions of internal versus external control as measured by the Rotter scale. Journal of Consulting and Clinical Psychology, 34, 226-228.

O'Brien, G. E. & Kabanoff, B. (1981). Australian norms and factor analysis of Rotter's Internal-External Control scale. Australian Psychologist, 16, 184-202.

Olson, G. H. (1983). Covariance structure analysis involving measured variables assumed to have higher-order factor structures. Paper presented at the Annual Meeting of the American Education Research Association, Division D, Montreal, Canada.

Pedhauzur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Prociuk, T. J. (1977). Concerning the evidence for a general factor in the Internal-External Control scale: A reanalysis of Franklin's Data. Journal of Consulting and Clinical Psychology, 45, 1199.

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80 (1, Whole No. 609).

Rotter, J. B. (1975). Some problems and misconceptions related to the construct of internal versus external control of reinforcement. Journal of Consulting and Clinical Psychology, 43, 56-67.

Sanger, S.S., & Alker, H.A. (1972). Dimensions of internal-external locus of control and the women's liberation movement. Journal of Social Issues, 28, 115-129.

Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Shavelson, R. J., & Marsh, H. W. (In Press). On the structure of self-concept. In R. Schwarzer (Ed.), Anxiety and cognitions. N. J. Erlbaum.

Stipek, D. J., & Weisz, J. R. (1981). Perceived personal control and academic freedom. Review of Educational Research, 51, 101-137.

Tanaka, J. S., & Huba, G. J. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. Journal of Personality and Social Psychology, 46, 621-635.

Tobacyk, J. (1978). Factor structure of Rotter's I-E scale in female Polish university students. Journal of Social Psychology, 106, 3-10.

Viney, (1974). The multidimensionality of perceived locus of control: Two replications. Journal of Consulting and Clinical Psychology, 42, 463-464.

Watson, J. M. (1981). A note on the dimensionality of the Rotter Locus of Control scale. Australian Journal of Psychology, 33, 319-330.

Wolk, S. & Hardy, R. (1975). The identifiability and consistency of the factor structure of locus of control. The Journal of Psychology, 89, 149-158.

Zuckerman, M. & Gerbasi, K.C. (1977). Dimensions of the I-E scale and the relationship to other personality measures. Educational and Psychological Measurment, 37, 159-175.

Table 1

The Number of Factors in the Rotter I-E Scale

Criterion	Time 1	Time 2
Number of eigenvalues[a] greater than 1.0	9	7
Number of statistically significant factors as determined by a chi-square test	4	5
Cattell's (1966) Scree Test[b]	4	7

a -- the 23 eigenvalues were: 3.31, 1.61, 1.38, 1.24, 1.92, 1.14, 1.10 1.05, 1.03, 0.98, 0.93, 0.88, 08.6, 0.80, 0.75, 0.70, 0.67, 0.66, 0.63, 0.59, 0.52, 0.50, and 0.43 for time 1; and 4.15, 2.00, 1.38, 1.3, 1.14, 1.11, 1.09, 1.0, 0.94, 0.87, 0.81, 0.78, 0.74, 0.72, 0.69, 0.66, 0.62, 0.57, 0.54, 0.50, 0.48, 0.46, 0.41.

b -- This test requires the researcher to plot a graph of the factor number against the eigenvalues, and to determine the the number of factors beyond which the plot levels off to form a relatively straight line that is nearly horizontal. However, as is the case in this application, the determination of the precise point where this occurs is somewhat subjective and different interpretations of the same plot are possible (Kim & Mueller, 1978).

Table 2

Summary of Results From Previous Factor Analyses of the Rotter Scale

Study	Percent Variance Explained[a] By Factors						Content of the Identified Factors						Number of Eigenvalues Greater Than 1.0
	1st	2nd	3rd	4th	5th	6th	GL	PC	SV	IC	AS	OS	
1 A	18%	9%	7%				x	x		x			---
B	16%	10%	8%				x	x		x			---
2 A	19%	10%	6%				x	x					5+
B	16%	9%	8%				x	x					5+
3 A	14%	9%	6%				x	x		x?			---
B	14%	8%	8%					x		x?	x?		---
4 A	13%	7%	7%	6%	6%		x	x	x?	x	x		5+
B	18%	7%	7%	6%	6%		x	x	x?	x	x		5+
5 A	16%	8%					x	x					---
B	20%	10%					x	x					---
6 A[b]							x?	x	x?	x	x		5+
7 A	11%	9%					x	x					---
B	12%	7%					x	x					---
8 A	13%	7%	7%	5%	5%	5%	x	x	x	x	x	x	6
9 A	8%	5%											---
B	12%	7%					x	x					---
10 A	6%	5%	5%	5%	4%		x	x	x?	x	x		5
11 A[c]	9%	8%	5%				x?	x?		x?			---
B	7%	10%	7%					x?					---
C	12%	9%	6%										---

Descriptions of Abbreviations Used to Characterize Factors.

GL = General luck, general control, fate.

PC = Political control.

SV = Success via personal initiative, personal control.

IC = Interpersonal control, likability, social relations.

AS = Academic situations.

OS = Occupational situations.

? = when presented with one of the above indicates that the classification of the factor is questionable.

Table 2 Continued

Study and Sample Descriptions:

1 = Abrahamson, Schludermann & Schludermann (1973): A = 120 university males; B = 113 university females.

2 = Cherlin and Bourque (1974): A = 161 university students; B = 100 adults from a general population. Eigenvalues were presented for 6 factors, but the interpretted solution was based on three factors.

3 = Dixon, McKee & McRae (1976): A = 98 male university students; B = 123 female university students.

4 = Garza & Widlak (1977): A = 244 Chicano undergraduate students; B = 203 Anglo undergraduate students.

5 = Lange and Tiggemann (1981): A = 277 first-year university students; B = 93 of the original sample retested 26 months latter.

6 = Little (1977): A = 418 undergraduates.

7 = Mirels (1970): A = 159 male undergraduates; B = 157 female undergraduates.

8 = O'Brien & Kabanoff (1981): A = 1921 individuals in a normative sample. Factor analyses were conducted on responses by subgroups in this study, but these results were not presented in sufficient detail to be presented.

9 = Viney (1974): A = 159 male adolescents; B = 134 female adolescents.

10= Watson (1981): A = 147 university students.

11= Wolk & Hardy (1975): A = 115 Black nursing students; B = 125 White nursing students; C = White education students. The authors did not attempt to label the empirically identified factors.

a -- Whenever possible, the percentage of variance is based on the variance explained by a factor before rotation, but some studies either reported percentage of variance explained after rotation or did not indicate whether the reported value was obtained before or after rotation. Since rotation procedures tend to distribute the explained variance more evenly than do the unrotated factors, this may produce a slight underestimate of the percentage of variance that can be explained by a single factor. In determining variance percentages and the number of eigenvalues greater than 1.0, a factor that accounts for 4.3% of the variance (i.e., 1/23 x 100% where 23 is the number of items in the Rotter scale) has an eigenvalue of 1.0. Thus the equation eigenvalue = variance explained/4.3% was used to compute some values that were not reported.

b -- the study did not report the percentage of variance explained by the unrotated factors, but Watson (1981) reported that the four and five factor solutions explained 38% and 44.5% of the variance.

c -- Factor loadings for three factors were presented for each sample, but neither the author of the original study nor we were able to make unambiguous interpretations of the content of those factors.

Note: Studies included in this table were limited to published factor analyses of the original Rotter scale. Not included in this table were the results of studies where: a) the results are unpublished; b) the actual factor analysis was not based on the 23 Rotter items; c) the way in which subjects responded to the items was altered d) the Rotter items were translated into a foreign language; or e) the results were not presented in sufficient detail to determine the content of identified factors and the percent of variance explained by these factors.

Table 3

Goodness-of-fit Indices for alternative models

Model	df	X^2 T1	X^2 T2	X^2/df T1	X^2/df T2	RMS T1	RMS T2	BBI T1	BBI T2	TLI T1	TLI T2
0	253	854	1307	3.38	5.17	.025	.032	---	---	---	---
1	230	392	583	1.70	2.53	.012	.016	.541	.554	.497	.511
2A	229	313	384	1.37	1.68	.010	.012	.633	.706	.594	.675
2B	227	306	383	1.35	1.69	.010	.012	.642	.707	.601	.674
3A	227	310	374	1.37	1.65	.010	.011	.637	.714	.595	.681
3B	216	294	347	1.36	1.61	.010	.011	.656	.734	.597	.689
4A	224	302	345	1.35	1.54	.010	.011	.646	.736	.601	.702
4B	207	267	287	1.29	1.39	.009	.009	.688	.780	.618	.732
4C	203	247	252	1.22	1.24	.009	.009	.711	.807	.640	.756
5A	220	290	336	1.32	1.53	.010	.011	.660	.743	.609	.704
5B	202	250	272	1.24	1.35	.009	.009	.707	.792	.633	.739
5C	201	241	252	1.20	1.25	.009	.009	.718	.807	.645	.758

$X2/df$ = The ratio of the chi-square to the degrees-of-freedom (df)

RMS = The Root Mean Square residual is a measure ot the average of residual variances and covariances for the original measured variables (Joreskog & Sorbom, 1981, p. I.41)

TLI = The Tucker-Lewis index is: $1-[(X^2_n/df_n)/(X^2_m/df_m)]$. The X^2/df_n is the chi-square/df ratio for the null model while the X^2/df_m is the corresponding value for the model being tested (Bentler & Bonett, 1980).

BBI = The Bentler-Bonett Index is: $1-[(X^2_n)/(X^2_m)]$. The X^2_n and X^2_m are the chi-square values for the null and tested models.

Note: Each model is defined by a number and a letter. The number refers to the number of factors hypothesized by the model. The letter refers to the structure: A indicates a simple structure where each item is allowed to load on only one factor; B indicates a complex structure in which some items load on more than one factor; C indicates a complex structure (as in B) but in addition some of the error/uniquenesses in THETA EPSILON are hypothesized to be correlated. Model 0, the null model, hypothesizes that all the measured variables are uncorrelated, and it is used primarily in defining the lower bound for the TLI and BBI. Model 5C differs from model 5B only in that the two items referring to occupation (items 9 & 13) are allowed to have correlated error/uniquenesses in THETA EPSILON.

Table 4

Four Tests of Factorial Invariance for Models 1 and 5C

Model	Test	Parameters Constrained To Be Equal	df	X^2	X^2/df	TLI	BBI	TC
0			506	2161	4.271	---	---	---
1	1	LY, PS, TE[a]	506	1086	2.15	.497	.498	.907
	2	LY, PS	483	1000	2.07	.537	.515	.980
	3	LY	482	997	2.07	.538	.515	.982
	4	none	460	976	2.12	.548	.503	1.000
5C	1	LY, PS, TE	477	654	1.37	.697	.679	.904
	2	LY, PS	453	561	1.24	.740	.710	.960
	3	LY	438	540	1.23	.750	.711	.972
	4	none	402	494	1.23	.771	.712	1.000

a -- LY = LAMBDA Y (factor loadings), PS = PSI (factor variances and covariances), TE = THETA EPSILON (error/uniquenesses)

b -- TC = Target Coefficient. The TC, a measure of a constrained model to explain variance accounted for the corresponding unconstrained model, is: $1 - [(X^2_{uc})/(X^2_{c})]$ where X^2_{uc} and X^2_{c} stand for the chi-squares for the model with no invariance constraints and the constrained model in which some elements are constrained to be equal across the multiple groups.

NOTE: For model 1 the chi-square value for test 1 differed significantly ($p < .01$) from the values for tests 2, 3 and 4, but no other differences were statistically significant. For model 5C the chi-square for test 1 was significantly higher than for the other tests, while the chi-square for test 4 was significantly smaller; tests 2 and 3 did not differ significantly from each other. Nevertheless, the goodness-of-fit indices for both models provide support for the invariance of factor loadings, and for factor variances and covariances, though perhaps not the error/uniquenesses.

BEST COPY AVAILABLE

Table 5

Confirmatory Factor Analysis of Responses to the Rotter I-E Scale:

The Ratio of Parameter Estimates to Standard Errors for Model 5C

Item	Factor Loadings (LAMBDA Y)[a] GL	PC	SV	IC	AS	Error/Uniqueness (THETA EPSILON)[b]
1	00	1.52/.33	1.49/.53	00	00	.19/.01
2	-0.71/.12	00	00	.10/.10	00	.17/.01
3	00	00	.49/.17	-.42/.14	00	.23/.01
4	.42/.17	-0.26/.17	00	00	-1.70/.65	.20/.02
5	00	.19/.24	-0.06/.32	.48/.18	00	.19/.01
6	00	00	00	1.00	00	.19/.02
7	00	.67/.19	.06/.29	00	00	.21/.01
8	00	00	00	00	-1.51/.31	.11/.01
9	00	-0.76/.18	-0.10/.27	00	00	.18/.10
10	-1.62/.19	00	00	00	00	.14/.01
11	00	-0.41/.18	.60/.29	00	00	.15/.01
12	00	-0.48/.16	.59/.26	00	00	.10/.01
13	00	.33/.09	00	.20/.11	00	.11/.01
14	1.72/.23	00	.19/.19	00	00	.14/.01
15	00	1.00	00	00	00	.17/.01
16	00	00	-0.07/.18	.74/.17	00	.20/.01
17	00	2.07/.55	2.54/.91	00	00	.15/.02
18	-1.51/.20	00	-0.07/.17	00	00	.14/.01
19	00	00	00	00	1.00	.07/.01
20	.16/.13	1.12/.12	00	00	00	.15/.01
21	00	00	-0.23/.19	-0.79/.18	00	.22/.01
22	00	00	1.00	00	00	.15/.01
23	1.00	00	00	00	00	.19/.01

Factor Variances & Covariances (PSI; Factor Correlations Above Diagonal)

	GL	PC	SV	IC	AS
GL	.038/.008	.494	-.478	.206	.376
PC	.025/.004	.066/.010	-.826	.507	.536
SV	-.016/.003	-.037/.006	.031/.007	-.270	-.516
IC	.010/.004	.031/.006	-.011/.005	.057/.014	.323
AS	.008/.002	.015/.003	-.010/.002	.008/.003	.011/.003

a -- see Table 1 for the definitions of the factors.

b -- these values represent the diagonal of the THETA EPSILON matrix. In addition, one off-diagonal value was estimated to represent the correlation between the two occupational items (9 & 13). This parameter estimate was -.028 with a standard error of .006.

NOTE: The parameter estimates are are from responses to time 1 and to time 2 when all parameter estimates were constrained to be invariant across both sets of responses. All items are scored as they originally appeared on the Rotter scale in that none were reflected, and the orientation of each factor is an arbitrary function of the orientation of the reference variable. Factor correlations, presented above the main diagonal in PSI, were not used in tests of the model, but are presented to aid in the interpretation of the findings.

Appendix I

A Priori Definition of Models With Simple and Complex Structures That Hypothesized 1 to 5 Factors To Explain Response to Rotter Scale.

Model	Factors	Items In Simple & Complex Structure	Items In Complex Structure Only
1	General I-E	all items	--
2A & B	General Luck	1,3,4,5,6,7,8,9,11,12 13,15*,16,17,19,20,21,22	--
	Political Control	2,10,14,18,23*	4,20
3A & B	General Luck	1,4,12,13,15*,17,20	5,7,9,11
	Political Control	2,10,14,18,23*	4,20
	Success Via PI[a]	3,5,6,7,8,9,11,16, 19,21,22*	1,12,14,17,18
4A & B	General Luck	1,4,12,13,15*,17,20	5,7,9,11
	Political Control	2,10,14,18,23*	4,20
	Success Via PI	5,7,8,9,11,19,22*	1,3,12,14,16,17, 18,21
	Interpersonal Rel	3,6*,16,21	2,5,13
5A, B, & C	General Luck	1,12,13,15*,17,20	4,5,7,9,11
	Political Control	2,10,14,18,23*	4,20
	Success Via PI	5,7,9,11,22*	1,3,8,12,14,16,17, 18,19,21
	Interpersonal Rel	3,6*,16,21	2,5,13
	Academic Situation	4,9,19*	---

a -- Success Via PI = Success Via Personal Initiative

* Indicates variable chosen to be the reference variable for each factor. It was chosen to be the item that best represented that factor and was most uniquely related to it on the basis of the a priori classification of items by the four raters as described earlier.

Note: Item numbers, 1 to 23, refer to the Rotter items after excluding the filler items. The definition of each factor is based on the classification of items into categories as described earlier. For models that specified 2 - 5 factors simple and complex structures were tested. For simple structures each item was allowed to load on only one factor (indicated in the column labelled "Items in Simple & Complex Structure") and for complex structures some items were allowed to load on more than one factor (indicated in the column labelled "Items in Complex Structure Only").