JOURNAL OF RESEARCH IN PERSONALITY **27**, 154–169 (1993)

Psychometric Properties of the Hope Scale: A Confirmatory Factor Analysis

MICHAEL A. BABYAK, C. R. SNYDER, AND LAUREN YOSHINOBU

University of Kansas, Lawrence

Confirmatory factor analysis was employed to test several psychometric hypotheses regarding the Hope Scale. Across four large samples of college students, a two-factor (agency and pathways) model of hope reproduced the observed data consistently better than did a one-factor model. Support also was found for the tenability of a higher-order latent construct overarching these two factors. Neither the assumption of parallel nor tau-equivalent measures were met, however, suggesting that the items within a given factor are not interchangeable. Reliability estimates of (1) the items as indicators of the first-order construct, and (2) the first-order constructs as indicators of the higher-order latent variable also are presented. Support for the higher-order two-factor model of hope also was found for both men and women. Some support also emerged for invariance of first- and second-order loadings, though this was not a consistent finding. Implications for the quantitative application of the Hope Scale are discussed.

An essential step in the development of a psychometric instrument is a thorough examination of the relations between the items on the instrument and the underlying constructs they are purported to measure. Traditionally, this task has been effected via exploratory factor analysis (EFA). An epistemologically more sound evaluation, however, is available via confirmatory factor analysis (CFA), within which the statistical fit of a hypothesized model is compared with the fit of theoretically important alternative models (Bollen, 1989; Joreskog & Sorbom, 1988). In the present paper we use CFA to evaluate the tenability of the proposed factor structure of a recently developed individual differences measure of psychological hope (Snyder, Harris, Anderson, Holleran, Irving, Sigmon, Yoshinobu, Gibb, Langelle, & Harney, 1991).

Snyder *et al.* devised a theory of hope based on the cognitive appraisal

---

Requests for reprints should be addressed to C. R. Snyder, 305 Fraser Hall, Doctoral and Postdoctoral Clinical Psychology Programs, Department of Psychology, The University of Kansas, Lawrence, Kansas, 66045.

0092-6566/93 $5.00

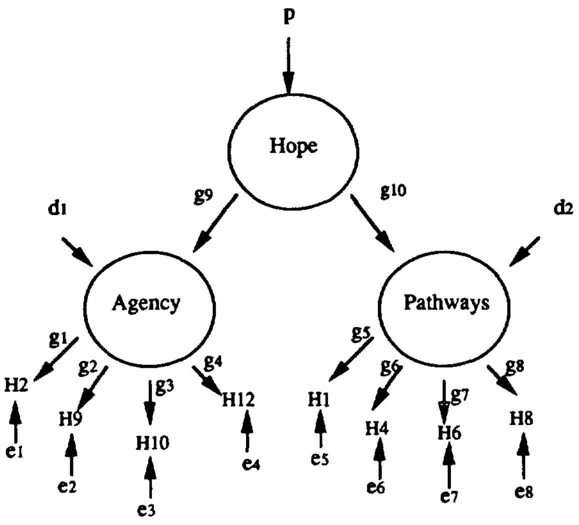

FIG. 1. Hypothesized higher-order model of hope.

that a person makes in relation to goal-related activities. Within the overarching cognitive set of hope are two goal-appraisal components, *agency* and *pathways*. The agency component is characterized by the willful sense of determination and energy to meet goals. The pathways component reflects an individual's perceived availability of ways to attain a goal. Elevated hope, as derived by the combined agency and pathways, is positively related to psychological adaptation. Using this model of hope, Snyder *et al.* developed and validated a brief self-report measure entitled the Hope Scale. Among the 12 items of the Hope Scale are 4 constructed to tap agency, 4 other items targeting pathways, and 4 distractive items to make the overall intent of the items less obvious. The instrument has adequate discriminant and convergent validity, and good discriminant utility in the prediction of coping styles and overall psychological well-being beyond variance attributable to other related constructs (e.g., optimism, positive and negative affectivity, locus of control, etc.; Snyder *et al.*, (1991).

In factor-analytic terms, agency and pathways are first-order latent variables "driven" by the second-order latent variable, hope. Demonstrating that the Hope Scale is consonant with a higher-order factor structure is of importance with respect not only to the theoretical assumptions of the Hope Scale, but to the manner in which the Hope Scale is implemented in analyses (Carver, 1989). Figure 1 depicts the relations among observed, first-order latent, and second-order latent variables as hypothesized by

Snyder *et al.* The observed variables, H2, H9, etc., represent items from the Hope Scale. The parameters g_1 through g_8 represent first-order factor loadings, and g_9 and g_{10} represent second-order loadings. The value e represents the error or uniqueness terms for the first-order solution; d values are error or uniqueness terms for the second-order solution. The term p represents the standardized variance of the second-order factor.

The present analyses address several hypotheses concerning the Hope Scale. Foremost, the fit of the two-factor hope model is compared with that of alternative models. The first of the comparison models is a "null" model in which the a priori specification is made that all observed variables are unrelated, i.e., the items on the scale have no loadings on any factors. Technically speaking, this reflects a restricting to zero of all covariances among the observed variables and allowing only the variances to be estimated. Although theoretically uninteresting, this is an important first test of any factor structure. Most importantly, the fit of null model can be used as a baseline against which alternative (and more theoretically interesting) models may be compared (Bentler & Bonett, 1980; Bollen, 1989). A second, alternative, model posits that all items on the Hope Scale load on one latent factor. In sum, three models will be tested: a baseline null model, a one-factor model, and the hypothesized two-factor model. If the instrument is consistent with Snyder's theory of hope, the two-factor model should reproduce the observed data significantly better than the alternative models.

As in EFA, CFA provides estimates of factor loadings. All first- and second-order loadings are expected to be equivalent in sign and substantial in magnitude. In addition to these tests, and contingent upon a satisfactory fit of the two-factor model, we will perform the additional test of tau-equivalence and parallelism of measures (see Dwyer, 1983; Joreskog & Sorbom, 1988). The tau-equivalent model constrains the loadings (g), and therefore the true-score variances (g^2), of the scale items on a given factor to be equivalent. The parallel model posits that both the true score variances (g^2) and the error variances (e) for the items associated with a factor are assumed equivalent. These assumptions are tested hierarchically: if the tau-equivalence assumption cannot be rejected, then the more restrictive parallel model may be evaluated. Although it is not crucial that the Hope Scale meet these criteria, the above measurement characteristics can influence the manner in which the scale is scored and implemented (Paunonen & Gardner, 1991; Perloff & Persons, 1988).

Finally, CFA provides descriptive information concerning the reliabilities of the constructs. Subtracting the squared value of the uniqueness term from unity ($1-d^2$) yields an estimate of the reliability of each first-order factor, which is also equivalent to the squared multiple correlation (R^2) coefficient for a given factor. CFA generated reliability estimates reflect the proportion of the variance of an observed variable accounted for by any systematic variance operating among the error terms in addition to that accounted for by the latent variable (Bollen, 1989).

In addition to the above analyses, the second order two-factor model of the Hope Scale, if supported, will be evaluated for equivalence across gender. Although frequently ignored, such a test is vital to comparisons of any construct across populations (Byrne, 1992; Drasgow & Kanfer, 1985). For example, it is not meaningful to talk about gender differences in the mean level of hope, or in the relation between hope and, say, depressed mood, if the Hope Scale is not measuring the same factors for men and women. For a construct to be considered psychometrically equivalent across populations, at the very least, the equivalence of structure and at least one corresponding loading should be demonstrated (Byrne, Shavelson, & Muthen, 1989).

METHOD

Subjects

Four cross-sectional data sets were available from the testing of University of Kansas undergraduate men and women taking a first-year psychology course. The data were collected in consecutive Fall semesters from 1987, 1988, 1989, and 1990, with sample sizes of 955, 472, 630, and 696, respectively.

Instrument

The Hope Scale contains 12 items, 4 of which are constructed to measure agency, and 4 others measure the pathways component; 4 additional items are distracters. Each of the 12 items consists of a statement and a Likert-type scale that ranges from 1 (agree) to 4 (disagree). The Hope Scale appears in Appendix 1.

Procedure

Analyses were first conducted for each sample collapsed across gender using the EQS program developed by Bentler (1989). Prior to factor analysis, the data were inspected for serious violations of distribution and variance assumptions via PRELIS (SPSS, 1990) a program designed specifically for this task. Parameters were estimated via maximum likelihood (ML) using the variance–covariance matrix generated by PRELIS as input. Reliability estimates were obtained by using the correlation matrix as input. In order to fix the metric and to increase the likelihood that the models were identified, one first-order loading per factor (g) was fixed to equal 1 and the second-order loadings were constrained to be equivalent. The variance of the second-order factor (p) was fixed to a value of 1 in order to set the scale for the second-order loadings (g).

Correlated errors of measurement were added post hoc to the models sample by sample based upon the Lagrangian Multiplier Test (Bollen, 1989) performed upon the hypothesized two-factor model shown in Fig. 1. The same pattern of correlated errors was specified in all remaining models within a given sample. These parameters represent systematic influences upon the observed item scores that are not predicted by theory and are therefore not specified a priori. Although they are added to the model post hoc, their inclusion in the models increases the likelihood that estimates of factor loadings, and in turn, reliabilities, will be unbiased (Bollen, 1989). Because correlated errors were added post hoc on a sample by sample basis, final models may differ in degrees of freedom across samples.

Because the traditional chi-square fit index and its associated probability are sensitive to departures from the assumption of multivariate normality and also are distorted upward by large sample sizes (Bentler & Bonett, 1980; Mulaik, James, Van Alstine, Bennet, Lind, & Stilwell, 1989), a number of additional fit indices are presented. The normed fit index (NFI) reflects the degree to which the hypothesized model can reproduce the observed variance/covariance matrix with respect to the null model. The nonnormed fit index (NNFI) and comparative fit index (CFI) also compare the null model to the hypothesized model. The values of the NFI, NNFI, and CFI generally range from zero to one, with values close to one reflecting a good fit (the NNFI and CFI can exceed this range under certain conditions). The NNFI and CFI are particularly attractive because they take into consideration sample size and degrees of freedom and behave better than the NFI when their values approach one (Bentler, 1990). In addition, the average absolute standardized residual (AASR) is presented as an index of fit. This value is a direct measure of the discrepancy between the observed and model reproduced variance–covariance matrices, the values of which approach one as fit improves. The chi-square difference test is used in the present case to evaluate the two-factor, tau-equivalence, and parallelism hypotheses, which are nested within the one-factor model depicted in Fig. 1. The difference between chi-square values of any two nested models evaluates the decrement in fit associated with a new set of parameter restrictions. A significant difference of chi-square value is interpreted as a significant worsening of fit of the model.

Tests of measurement invariance across gender were conducted sample by sample in the following manner. First, the a priori two-factor model was estimated separately for men and women with all correlations among error terms set to zero. A post hoc Lagrangian Multiplier Test was used as a guide to freeing correlated errors until a satisfactor fit was achieved without creating identification problems. These best fitting models were then estimated for men and women simultaneously with no invariance restrictions imposed across groups. This procedure yields a single chi-square value and set of fit indices that reflect the combined fit of the models for men and women. Gender differences were tested hierarchically from least to most restrictive models, beginning with the hypothesis of equivalence of factor structure, followed by first-order loadings, second-order loadings, and unique variance terms associated with the first-order factors and scale items (Bentler, 1989). Since each model is nested within the previous model, the difference of chi-square test also can be applied here.

RESULTS

The distribution of each scale item exhibited a small but consistent negative skew, suggesting that respondents tended to endorse values toward elevated agency and pathways. Table 1 displays the fit of the one- and two-factor models before the addition of correlated errors (marked "no c.e." in table). Under this specification, the two-factor model yielded a poor fit by the chi-square criterion, but an adequate fit by the other fit indices. The chi-square difference test between these two models suggests that the two-factor model reproduces the observed data significantly better than does the one-factor model. After the addition of correlated errors, Table 1 shows that for Samples 2 and 3, the two-factor model fit adequately by the chi-square criterion at the .05 level; Sample 1 at the .01 level. This model was rejected in Sample 4 according to the chi-square criterion.

TABLE 1

FIT INDICES FOR HYPOTHESIZED AND COMPETING MODELS[a]

	df	χ^2	*p*	NFI	NNFI	CFI	AASR	χ^2_{diff}	*p*
Sample 1: *N* = 955									
1-factor (no c.e.)	20	214.76	<.001	.855	.812	.866	.046		
2-factor (no c.e.)	19	105.43	<.001	.924	.912	.940	.032	109.33(1*df*)	<.001
1-factor (with c.e.)	16	108.87	<.001	.92	.888	.936	.031		
2-factor (with c.e.)	15	31.00	<.010	.979	.979	.989	.014	77.87(1*df*)	<.001
No higher-order	16	314.64	<.001	.787	.640	.794	.107	283.64(1*df*)	<.001
Tau-equivalent	19	56.18	<.001	.962	.962	.974	.035	25.18(4*df*)	<.001
Parallel	25	94.01	<.001	.936	.947	.952	.041	37.83(6*df*)	<.001
Sample 2: *N* = 472									
1-factor (no c.e.)	20	166.61	<.001	.780	.719	.799	.055		
2-factor (no c.e.)	19	82.23	<.001	.892	.872	.913	.040	83.37(1*df*)	<.001
1-factor (with c.e.)	14	97.01	<.001	.871	.771	.886	.042		
2-factor (with c.e.)	13	20.78	.078	.972	.977	.989	.018	76.23(1*df*)	<.001
No higher-order	14	91.64	<.001	.879	.787	.894	.096	70.86(1*df*)	<.001
Tau-equivalent	17	43.47	<.001	.942	.940	.964	.035	22.69(4*df*)	<.001
Parallel	23	131.29	<.001	.826	.818	.851	.064	87.82(6*df*)	<.001
Sample 3: *N* = 630									
1-factor (no c.e.)	20	203.05	<.001	.809	.752	.823	.052		
2-factor (no c.e.)	19	104.55	<.001	.901	.878	.917	.039	98.49(1*df*)	<.001
1-factor (with c.e.)	15	87.43	<.001	.918	.870	.930	.036		
2-factor (with c.e.)	14	15.49	.346	.985	.997	.999	.012	71.94(1*df*)	<.001
No higher-order	15	176.84	<.001	.833	.708	.843	.037	227.31(1*df*)	<.001
Tau-equivalent	18	48.24	<.001	.955	.955	.971	.050	32.79(4*df*)	<.001
Parallel	24	61.96	<.001	.942	.957	.963	.040	13.68(6*df*)	<.05
Sample 4: *N* = 684									
1-factor (no c.e.)	20	161.13	<.001	.904	.881	.915	.037		
2-factor (no c.e.)	19	109.23	<.001	.935	.920	.945	.028	51.89(1*df*)	<.001
1-factor (with c.e.)	15	67.17	<.001	.960	.941	.968	.024		
2-factor (with c.e.)	14	35.51	<.001	.979	.974	.974	.002	69.29(1*df*)	<.001
No higher-order	15	375.81	<.001	.773	.593	.782	.169	414.98(1*df*)	<.001
Tau-equivalent	18	40.19	<.001	.976	.979	.987	.023	19.97(4*df*)	<.001
Parallel	24	93.70	<.001	.944	.951	.958	.033	47.45(6*df*)	<.001

[a] See text for explanation of headings.

The four additional fit indices, however, indicate that the two-factor model is probably a very well-fitting model in all samples and that the chi-square value may be quite inflated due to sample size. In each sample, the chi-square difference and each additional fit index suggests that the two-factor model reproduced the observed data significantly better than did the one-factor model. Also included in Table 1 are the results of the formal test of the tenability of the higher factor, in which the higher order two-factor model with correlated errors is compared with the same model except that the second-order loadings are constrained to zero, i.e., no higher-order factor. The significantly poorer fit indices associated with the model that hypothesizes no higher-order factor clearly support the model with the higher-order factor.

No support for tau-equivalence was found, however, meaning that the loadings and therefore the true score variances of the items differed within a factor. In each case, the difference of chi-square between the congeneric and tau-equivalent models suggests a significant worsening of fit when the tau-equivalence restriction is imposed. As presented earlier, rejecting the tau-equivalent model renders the test of parallel measures meaningless; the tests of parallel measures are reported merely for the sake of completeness. Table 2 shows that across all samples the first- and second-order loadings were of the same sign (positive) and substantial, as predicted. Also presented in Table 2 are the correlated errors of measurement added to the models post hoc based upon the Lagrangian Multiplier Test.

Finally, the explained variance estimates for the individual scale items and for the first-order factors are reported as squared multiple correlations in Table 3. These values may be interpreted as the reliability of a scale item as an indicator of its associated latent construct, and the reliability of a first-order latent variable an indicator of its associated second-order latent variable, respectively. Reliabilities ranged from .26 to .86 for the scale items, and .96 to .99 for the agency and pathways dimensions.

Table 4 summarizes the results of the nested tests of measurement invariance across gender. The first hypothesis (H-form) posits that the form of the model is the same for both groups, i.e., that a second-order two-factor model adequately reproduces the observed data for both men and women. This hypothesis was not rejected in any of the samples. First- and second-order factor loadings were all positive and substantial for both men and women. Considering the chi-square difference test first, the hypothesis that the corresponding factor loadings were equivalent across gender (H-loadings1) was not rejected in Samples 1 and 3, but rejected in Samples 2 and 4. In Samples 1, 2, and 3, the hypotheses of invariant second order loadings (H-loadings2) was not rejected, but was rejected in sample 4. The invariance of the disturbance terms associated with the first-order factors (H-residual) was supported in all of the samples. The

TABLE 2

FIRST-ORDER LOADINGS OF OBSERVED VARIABLES ON AGENCY (H2 TO H12) AND PATHWAYS (H1 TO H8) AND SECOND-ORDER LOADINGS OF AGENCY AND PATHWAYS ON HOPE

	Sample 1	Sample 2	Sample 3	Sample 4
Hope 2[a]	1.000[b]	1.000	1.000	1.000
	.373[c]	.575	.425	.608
Hope 9	1.307	.877	1.075	1.095
	.445	.444	.412	.588
Hope 10	1.993	1.058	1.888	1.185
	.730	.589	.750	.677
Hope 12	1.819	.992	1.930	1.160
	.682	.595	.790	.690
Hope 1[a]	1.000	1.000	1.000	1.000
	.546	.429	.517	.677
Hope 4	1.384	1.696	1.603	.942
	.649	.633	.700	.588
Hope 6	1.316	1.749	1.345	1.046
	.691	.769	.677	.667
Hope 8	1.045	1.107	.889	1.100
	.672	.356	.534	.715
Agency	.215	.226	.209	.390
	.949	.896	.731	.902
Pathways	.215	.226	.209	.390
	.697	.627	.843	.949

Correlations among measurement errors

Sample 1	Sample 2	Sample 3	Sample 4
H1,H4 = .220	H1,H4 = .242	H1,H4 = .341	H1,H4 = .093
H1,H8 = .117	H2,H12 = .171	H2,H12 = .207	H1,H12 = −.179
H6,H8 = −.346	H2,H8 = .239	H6,H8 = −.549	H8,H9 = .133
H9,H12 = −.225	H1,H2 = .110	H1,H2 = .093	H9,H10 = .280
	H2,H6 = .204	H8,H9 = .152	H2,H6 = .147
	H8,H9 = .174		

[a] Values have been fixed to one in order to set metric.
[b] Unstandardized coefficient.
[c] Standardized coefficient.

hypothesis of invariant unique variance terms (H-errors) was not rejected in Sample 3, but rejected in all other samples. Somewhat in contrast to the chi-square difference tests, however, nearly all of the above tests of invariance yielded relatively satisfactory fit indices. In only one case, the most restrictive hypothesis in Sample 1, does any fit index fall below .900. Finally, the unstandardized and standardized loadings for men and women appear in Table 5. As expected, all loadings were substantial and positive for both groups.

TABLE 3

Reliability (R^2) Estimates for Items and First-Order Factors

	Sample 1	Sample 2	Sample 3	Sample 4
Hope 2	.26	.55	.33	.60
Hope 9	.36	.36	.31	.57
Hope 10	.78	.58	.81	.71
Hope 12	.71	.59	.86	.73
Hope 1	.51	.35	.81	.71
Hope 4	.67	.65	.74	.57
Hope 6	.73	.83	.71	.69
Hope 8	.70	.54	.49	.69
Agency	.99	.99	.98	.97
Pathways	.97	.96	.99	.99

TABLE 4

Fit Indices for Nested Invariance Hypotheses

	df	χ^2	*p*	NFI	NNFI	CFI	χ^2_{diff}	*p*
Sample 1: men = 447, women = 508								
H-form	34	65.21	<.001	.957	.965	.979		
H-loadings1	40	77.35	<.001	.949	.964	.975	12.14(6*df*)	NS
H-loadings2	41	80.521	<.001	.947	.963	.973	3.17(1*df*)	NS
H-residual	43	82.67	<.001	.946	.965	.963	2.15(2*df*)	NS
H-errors	51	105.95	<.001	.931	.959	.963	23.28(8*df*)	<.005
Sample 2: men = 241, women = 231								
H-form	29	44.19	.035	.945	.961	.980		
H-loadings1	35	61.49	.003	.924	.944	.965	17.30(6*df*)	<.01
H-loadings2	36	62.98	.004	.944	.944	.964	1.49(1*df*)	NS
H-residual	38	63.00	.006	.922	.951	.967	.03(2*df*)	<.05
H-errors	46	88.24	<.001	.891	.932	.944	25.23(8*df*)	<.005
Sample 3: men = 287, women = 343								
H-form	30	39.68	.112	.963	.982	.991		
H-loadings1	36	48.11	.085	.955	.982	.988	8.44(6*df*)	NS
H-loadings2	37	49.65	.080	.954	.981	.988	1.53(1*df*)	NS
H-residual	39	53.45	.062	.950	.980	.986	3.80(2*df*)	NS
H-errors	47	59.50	.042	.945	.985	.988	6.06(8*df*)	NS
Sample 4[a]: men = 333, women = 345								
H-form	29	44.70	.032	.951	.965	.982		
H-loadings1	35	68.37	<.001	.926	.938	.961	23.67(6*df*)	<.001
H-loadings2	36	79.69	<.001	.913	.921	.949	1.32(1*df*)	NS
H-residual	38	79.86	<.001	.913	.929	.952	.17(2*df*)	NS
H-errors	46	87.26	<.001	.905	.942	.952	7.46(8*df*)	NS

[a] Total *N* is different from combined groups analysis due to missing data.

TABLE 5

FIRST-ORDER LOADINGS OF OBSERVED VARIABLES ON AGENCY (H2 TO H12) AND PATHWAYS (H1 TO H8) AND SECOND-ORDER LOADINGS OF AGENCY AND PATHWAYS ON HOPE: MEN AND WOMEN

	Sample 1		Sample 2		Sample 3		Sample 4	
	Men	Women	Men	Women	Men	Women	Men	Women
Hope 2[a]	1.000[b]	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	.539[c]	.523	.444	.567	.540	.644	.594	.616
Hope 9	1.310	1.434	2.007	.831	1.492	1.022	.443	1.009
	.606	.646	.738	.430	.657	.573	.228	.597
Hope 10	1.430	1.303	1.792	.958	1.347	.847	.821	1.072
	.712	.690	.750	.606	.682	.548	.479	.618
Hope 12	.897	.884	1.174	.923	.868	.856	1.175	1.166
	.570	.545	.584	.661	.525	.656	.764	.721
Hope 1[a]	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	.391	.518	.508	.637	.518	.486	.479	.627
Hope 4	1.226	1.049	.978	.649	1.060	1.021	.968	1.240
	.428	.509	.454	.352	.497	.430	.392	.676
Hope 6	2.090	1.172	1.208	1.034	1.523	1.212	1.095	1.316
	.816	.577	.614	.612	.686	.564	.471	.781
Hope 8	1.986	1.134	1.523	.871	1.368	1.224	1.185	.975
	.789	.573	.825	.563	.663	.567	.564	.597
Agency	.223	.248	.221	.288	.254	.267	.218	.363
	.743	.820	.905	.806	.851	.750	.653	.905
Pathways	.223	.248	.221	.288	.254	.267	.218	.363
	.943	.793	.701	.717	.846	.972	.804	.875

[a] Values have been fixed to one in order to set metric.
[b] Unstandardized coefficient.
[c] Standardized coefficient.

DISCUSSION

The fit indices consistently favor the hypothesized higher-order two-factor hope model over the single-factor model and the model with no relation specified between agency and pathways, a result that is strengthened by replication over four large samples. It is important to point out that an adequate fit can never be construed as confirmation of a model; these results only constitute support that one model is more consistent with the observed data than competing models. As is always the case in CFA, a number of other models also will fit the data. Moreover, because the Hope Scale was specifically constructed to measure two factors based upon both a priori and a posteriori information, these analyses represent an evaluation of the consistency of the Hope Scale with theory rather than any test of the theory itself. In other words, one central question addressed by these analyses is whether the Hope Scale does what was intended, i.e., taps two factors rather than one.

With respect to the higher-order factor, we followed Gorsuch's (1983) recommendation that correlated factors be modeled as higher-order factors in order to extract from the data the best possible substantive understanding. Our interpretation of this suggestion is that if sufficient theoretical justification exists, some attempt should be made to explain the covariances among factors in substantive terms. In concept, this endeavor does not differ from attempting to model the covariances among the items via the first-order factors. We have demonstrated the tenability of this higher-order model by virtue of the presence of substantial second-order loadings of agency and pathways on hope, and the significant worsening of fit when these loadings are fixed to zero. Therefore, at least mathematically, some higher-order dimension could underlie pathways and agency. As alternatives to the higher-order model, we could have chosen to specify a model in which agency and pathways were reciprocal causes of one another, or as causes, rather than indicators, of the third latent construct hope (Bollen, 1987). When compared to these alternative models, however, the higher-order model clearly best reflects the proposed structure of hope as outlined in Snyder *et al.* (1991).

Snyder *et al.*'s conceptualization of hope also clearly specifies that agency and pathways are distinct but highly related entities, and that it is when these two components act in tandem that hope is operative and an effective predictor of criterion variables. In practice, then, using the scale as a unitary entity should yield the best predictive power with respect to criterion variables. To date, the majority of research conducted with the Hope Scale has indeed employed it as a unitary entity. We maintain, however, that this single variable that is being used in the research is the higher-order latent variable, not simply a unidimensional first-order construct. The psychometric evidence presented above supports this view. In practice, of course, this is a meaningless distinction. Nevertheless, in furthering the conceptual understanding of the hope theory, such information is crucial. As Carver (1989) suggests, in order for a multidimensional construct to be employed appropriately as a single dimension, it is necessary to demonstrate the existence of an underlying latent variable upon which the lower order dimensions converge. Carver also notes that the lower order dimensions should not be differentially associated with criterion variables. Previous regression analyses evaluating the separate effects of agency and pathways upon a criterion variable have yielded significant main effects, but no interactions between agency and pathways (see Snyder *et al.*, 1991). Therefore, available data suggest that the agency and pathways components are not differentially associated with outcomes, thus adding further support for the higher-order hypothesis.

Presenting the more complex conceptual structure also bears on a fundamental goal of research—the ability of future investigations to attempt to falsify and improve upon the theoretical assumptions posited by Snyder *et al.* (1991). By laying all of the conceptual cards on the table, that is, by being as explicit as we can about the conceptual structure of hope, we enrich the potential of future investigations of this theory. Future investigations, for example, might attempt to demonstrate the tenability of alternative factor structures that explain the observe data gathered from the Hope Scale. Other investigations might focus on the more specific lower-order dimensions as predictors, while others might be more interested in the more general higher-order construct. This latter exploration might further explicate the merits, or lack thereof, of using the combined agency and pathways scores as a single dimension.

Across all samples, the first-order loadings are, as expected, the same in sign and substantial. The loadings in general are moderately high. Therefore the hypothesized factors explain a moderate amount of variance of each item. The rejection of the tau-equivalence and parallel assumptions, however, suggests that the items within a factor are not equivalent in their relation with the underlying construct and therefore probably not interchangeable. Failure to meet these assumptions probably bears few consequences for the implementation of the Hope Scale, other than providing a more complete understanding of the scale. Some controversy exists over whether it is desirable to incorporate these weights when calculating total scale or subscale scores (e.g., using factor scores or the regression method) or to simply use unit weighting. The consensus in the literature appears to be that, under most circumstances, unit weighting, regardless of the "true" item weights, does not affect estimates of relations between the scale and other variables (e.g., Dawes & Corrigan, 1974; Pedhazur & Schmelkin, 1991; Wainer, 1978). Not all, however, agree with this conclusion. Perloff & Persons (1988), for example, argue that under many conditions, failure to appropriately weight items may bias the coefficient value and significance test of parameters that estimate the relation between a construct and some criterion.

Regardless of the interpretation of the weighted scores literature, a more satisfactory method for using the Hope Scale might be to employ structural modeling. Structural modeling techniques address any concerns that might arise regarding weighting, and more importantly, also incorporates information regarding the reliability of constructs into the estimates of structural relations between the latent construct and criterion variables. The latter feature is particularly important because the presence of measurement error has been associated with considerable bias in parameter estimates (Dwyer, 1983). In the case of the Hope Scale, the reliabilities of the agency and pathways constructs are quite good as indicators of the broader hope construct, but as is the case for many self-report instruments, considerable measurement error exists in the items as indicators of the first-order constructs. In view of this, the best method available to address any concerns regarding measurement error is to use a full structural equation, or latent variable, modeling technique. Having provided these caveats about the Hope Scale, we would hasten to add that similar concerns would probably arise if other existent individual differences measures were put to such new and stringent tests. Indeed, it should be noted that a number of commonly used psychometric instruments would be more appropriately utilized—and for similar reasons—in this manner.

The analyses of gender differences suggest that the hierarchical two-factor model is tenable for both men and women. By the chi-square difference criterion, the first-order factor loadings appear equivalent in two samples, but differ in two other samples. Despite these significant chi-square difference tests, however, the invariant first-order loading restriction still yields relatively good fit in all samples according to the other indices. The same is true for the second order equivalence restriction that failed to meet the chi-square criterion in Sample 4. The equivalence of the disturbance terms associated with the first-order factor was supported in all samples, but the uniqueness terms associated with the items are clearly different across genders. Differences in these latter terms, however, are relatively unimportant with respect to demonstrating the equivalence of a construct across groups (Bentler, 1989).

The overall picture of these invariance hypotheses then is consistent with earlier exploratory factor analyses of the Hope Scale (Snyder *et al.*, 1991); the proposed hierarchical two-factor model is tenable for both men and women and the items seem to be measuring the same factors in both groups. This can only be a tentative conclusion, however, because of the discrepancy between the chi-square difference tests and the other fit indices. Because of the large sample sizes, however, the chi-square difference test may be quite sensitive to even substantively negligible differences in the fit of the nested models. Hence, we favor the conclusion that the constructs are invariant at the level of the first- and second-order loadings. Although we favor the conclusion of invariance, it is interesting to note that the biggest gender differences in the loadings in the two samples in which this hypothesis was not maintained by the chi-square difference criterion are associated with the loadings on agency. However, women have lower loadings than do men on this construct in Sample 2, but higher loadings than do men in Sample 4. One interpretation of this trend is that agency is a less stable construct for women than for men. Given the current transitional nature of women's gender roles, it may be that women in this culture and cohort are more conflicted about agentic attitudes than are their male counterparts.

An application of the multi-trait multi-method technique combined with the structural equation model approach may be quite useful in further clarifying the invariance hypotheses and the group differences in the uniqueness terms. In this case, one might specify separate structural equation models for men and women. This would allow the parameter estimates in each group to incorporate group-specific information about measurement error, resulting in less biased parameter estimates (Dwyer, 1983), and therefore less biased evaluation of group differences. Alternatively, Muthén (1989) has suggested a method by which heterogeneity in construct equivalence may be dealt with by employing the MIMIC modeling technique.

One area of further concern is the interpretation of the correlated errors of measurement that were added post hoc to the models in the present analyses. In the combined groups analyses, the error terms associated with H1 and H4 are correlated in all four samples; in addition, there are several other correlated errors that are common to at least two of the samples. In the gender invariance analyses, the patterns of correlated errors seems less easily characterized, differing from sample to sample and across gender. Although correlated errors can be interpreted as the presence of a systematic influence (for example, method-specific variance) upon the item responses that exists outside the specified model, the source or nature of this influence cannot be determined with the present data. Comparing the fit of the models with and without correlated errors suggests that this systematic influence is not trivial, and warrants further explication. The correlated errors between items H1 and H4, for example, may reflect the similarity in the wording of these items, and may suggest that these items are redundant. Again, one way of clarifying this issue would be a multi-trait multi-method approach to evaluating the hope construct.

To summarize, the hypothesized higher-order two-factor hope model consistently reproduced the observed data better than the null and one-factor models. By at least this criterion then, the Hope Scale is successful in its putative function. Consonant with the hope theory, these findings suggest that although agency and pathways are highly related, they do not constitute a single factor, rather, they are relatively distinct entities that converge upon a broader latent construct. Like many self-report measures, however, the items of this scale are not interchangeable within their respective factors and contain measurement error. The proposed hierarchical two-factor model is tenable for both men and women. Support for the invariance of factor structure and corresponding factor loadings across gender also was found.

APPENDIX 1: THE HOPE SCALE

Directions: Read each item carefully. Using the scale shown below, please select the number that best describes YOU and put that number in the blank provided. Thank you.

1 = Definitely false 3 = Mostly true
2 = Mostly false 4 = Definitely true

__ 1. I can think of many ways to get out of a jam. (Pathways)
__ 2. I energetically pursue my goals (Agency)
__ 3. I feel tired most of the time. (Distracter)
__ 4. There are lots of ways around any problem. (Pathways)
__ 5. I am easily downed in an argument. (Distracter)
__ 6. I can think of many ways to get the things in life that are most important to me. (Pathways)
__ 7. I worry about my health. (Distracter)
__ 8. Even when others get discouraged, I know I can find a way to solve the problem. (Pathways)
__ 9. My past experiences have prepared me well for my future. (Agency)
__ 10. I've been pretty successful in life. (Agency)
__ 11. I usually find myself worring about something. (Distracter)
__ 12. I meet the goals that I set for myself. (Agency)

REFERENCES

Bentler, P. M. (1989). *EQS: Structural equations program manual.* Los Angeles, CA: BMDP Statistical Software, Inc.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin,* **2,** 238–246.

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

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Byrne, B. M. (1989). Multigroup comparisons and the assumption of equivalent construct validity across groups: Methodological and substantive issues. *Multivariate Behavioral Research,* **24,** 503–523.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin,* **105,** 456–466.

Carver, C. S. (1989). How should multifaceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology,* **56,** 577–585.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin,* **81,** 95–106.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology,* **70,** 662–680.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences.* New York: Oxford Univ. Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL.* Baltimore, MD: Johns Hopkins Univ. Press.

Joreskog, K. G., & Sorbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS Inc.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennet, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin,* **105,** 430–445.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika,* **54,** 557–585.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design & analysis: An integrated approach.* Hillsdale, NJ: Erlbaum.

Perloff, J. F., & Persons, J. B. (1988). Biases resulting from the use of indexes: An application to attributional style and depression. *Psychological Bulletin,* **103,** 95–104.

Paunonen, S. V., & Gardner, R. C. (1991). Biases resulting from the use of aggregated variables in psychology. *Psychological Bulletin,* **109,** 520–523.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., Gibb, J., Langelle, C., & Harney, P. (1991). The will and the ways: Development and validation of an individual differences measure of hope. *Journal of Personality and Social Psychology,* **60,** 570–585.

SPSS. (1990). *SPSS LISREL 7 and PRELIS: User's reference guide.* Chicago: SPSS Inc.

Wainer, H. (1978). On the sensitivity of regression and regressors. *Psychological Bulletin,* **83,** 267–273.