

The Attitudes Toward Forcible Date Rape (FDR) Scale: Development of a Measurement Model

Gloria J. Fischer¹ and Jieming Chen²

Accepted: January 10, 1994

Using data from 341 female and 237 male college students that were collected previously (Fischer, G. J., Archives of Sexual Behavior, 15, 457-466, 1986), several true score measurement models were used to describe the variance-covariance structure of a nine-item attitude toward a forcible date rape (FDR) scale. The congeneric true score model fit the data best, but not satisfactorily. By deleting "noncongeneric" items, a six-item, unweighted linear composite variable based on the congeneric true score model was shown to fit the data, and reliabilities and validities based on this model proved satisfactory for females, males, and the total sample. The factor structure of the model for females and for males was not equivalent, but the general patterns were similar. Suggestions for further research included a validation study of the six-item scale on an independent sample and a comparison of 5- and 7-point Likert response scales to see if the lack of model equivalence by gender could be due, in part, to more variability in attitudes toward forcible date rape in men than women.

KEY WORDS: date rape, forcible date rape scale, rape myth beliefs.

INTRODUCTION

Feild (1978) developed a 32-item attitudes toward rape scale consisting of items about the act, the victim, and the offender. Based on eigenvalues greater than 1.0 from a principal-components analysis, eight factors were extracted and labeled. Many of these items have been called rape

¹Washington State University, Pullman, Washington 99163.

²The University of Michigan, Ann Arbor, Michigan 48109.

myth beliefs (e.g., Costin, 1985). Costin (1985) used seven of Feild's (1978) items, reworked nine others, and added four. A principal-components analysis yielded three factors. The first, which loaded on eight items and accounted for 64% of the variance, appeared to be Feild's (1978) first factor, "women's responsibility for rape." Costin (1985) labeled the second factor (two items) "role of consent" and the third factor (five items) "rapist's role and motivation." However, none of the 20 items in Costin's (1985) scales refer to date rape or to behavior on a date.

Burt (1980) compiled a 14-item rape myth acceptance scale (RMA; Cronbach's $\alpha = .875$), but only 1 of the 14 items refers to behavior on a date. Burt (1980) related RMA scores to a 6-item "acceptance of interpersonal violence" scale (Cronbach's $\alpha = .586$), a 9-item "adversarial sexual beliefs" scale (Cronbach's $\alpha = .802$), a 10-item "sexual conservatism" scale (Cronbach's $\alpha = .811$), and a 9-item "sex role stereotyping" scale (Cronbach's $\alpha = .800$). Koralewski and Conger (1992) used all of Burt's (1980) scales but found that only scores on the attitudes toward interpersonal violence scale discriminated 14 college males who admitted having engaged in forced intercourse from 21 males who had engaged in lesser forms of sexually coercive activity and 29 who had not engaged in any sexually coercive acts. Again, none of the items on this scale involve interpersonal violence on a date.

Dull and Giacopassi (1987) attempted to develop an attitudes toward rape scale. A principal-components analysis of their 15-item scale showed two factors. They labeled the five items loading on one factor "attitudes toward sex and dating." Five other items loading on the second factor were labeled "attitudes toward rape and date rape." However, only two of the five items on the latter scale mention date rape. Neither scale reliabilities nor gender differences were addressed.

Date or acquaintance rape is very much less likely to be reported than stranger rape. When reported, charges are unlikely to be made, and if made, juries are unlikely to convict. This is all the more serious, since date or acquaintance rape is far more common than stranger rape. Given the prevalence of date or acquaintance rape, the development of an attitudes toward date rape scale and rigorous investigation based on measurement theory is desirable. As indicated previously, existing, commonly used scales do not address date rape and are not based on measurement theory. The purposes of the present study were to describe the development of an attitudes toward date rape scale, summarize data on its predictive utility, and address scale reliability and validity. Such a scale was developed initially by Giarusso, Johnson, Goodchilds, and Zellman (1979). They asked high-school students, "Is it all right if a male holds a female down and physically forces her to engage in sexual intercourse?" (yes or no), followed

by nine circumstances, such as “when a girl gets a guy sexually excited.” Results of a factor analysis of acceptable–not acceptable responses to the nine circumstances yielded a single factor that the authors labeled “force.” Force was found least acceptable in the circumstance “when a guy spends a lot of money on a girl” and most acceptable “when a girl gets a guy sexually excited.” Mahoney (1983) extended findings to college students. Instead of the “Is it all right...” question, however, he used a forcible date rape vignette, followed by a “Would you call this rape?” question and the nine circumstances. Following Mahoney (1983), Fischer (1986) scaled the originally dichotomous response to the “Would you call this rape?” question, added the question “Who is to blame? [the male, the situation, society, or the female” (M. Ward, personal communication, 1978)] and scaled the originally dichotomous response to the nine circumstances. The resulting scale (see Appendix) was named the attitudes toward forcible date rape (FDR) scale (Fischer, 1986, 1987) and is the sum of the scores on the nine circumstances, in which definitely acceptable is coded as 1, mildly acceptable as 2, and definitely unacceptable as 5.

The FDR scale has been shown to have predictive utility. Fischer (1986, 1987) found that college students with relatively more accepting attitudes toward forcible date rape (i.e., did not judge the male’s behavior “definitely unacceptable” in any of the nine circumstances) were less sure that the date rape really was rape, had more traditional attitudes toward women, were more self-sexually permissive, had less accurate sexual knowledge, and were more inclined to blame the situation or society than were students with less accepting attitudes. FDR also was found to vary as a function of gender (lower in males) and ethnicity [lower in Hispanics, especially males, than in Whites (Fischer, 1987)], where lower scores are more accepting of date rape. FDR was a significant predictor of sexually coercive college males [i.e., males who self-reported having told lies or made false promises to have sex (Fischer, 1992)].

Findings cited suggest predictive utility of the FDR scale. However, it has not been subjected to rigorous investigation (nor has any other attitudes toward rape scale) from a measurement theory approach. To do so, the present study addressed the following questions: (1) Is the FDR scale unidimensional, as alleged? (2) Does it measure attitudes toward forcible date rape? (3) Does it do so consistently? (4) Do all nine individual items comprising the scale contribute equally? and finally, (5) Does the scale vary by gender? Consistency (reliability) was assessed using Pearson correlations. The method used to address the evaluation of Question 1 regarding dimensionality was done using exploratory factor analysis. The method used to address Questions 2 and 4 was the fitting of classic true score models of the variance–covariance structure of the nine items, using

confirmatory factor analyses. To test Question 5, the similarity of the models for women and men was assessed by confirmatory factor analyses, by gender.

METHOD

Subjects. Data for 341 female and 237 male college students were taken from a study reported elsewhere (Fischer, 1986). The students, mostly sophomores and predominantly White, were volunteers from human sexuality and introductory psychology classes in Fall semesters, 1982 and 1983, and the students received class credit for participating. They took the FDR scale anonymously, in a survey of sexual attitudes, knowledge, and experience.

RESULTS

Dimensionality. As shown in the Appendix, each FDR scale item has a 5-point Likert response scale ranging from "definitely acceptable" (coded 1) to "definitely unacceptable" (coded 5). Univariate statistics and Pearson correlations of the nine items, by gender, are shown in Table I. Inspection of Table I reveals remarkable consistency of positive correlations among the nine items, consistent with the pattern of a single underlying latent variable influencing the items. Given this pattern and Giarusso and co-workers' (1979) finding of a single factor from their analysis, the conclusion that one single factor adequately describes the covariance structure seemed reasonable. However, uniformly high correlations among items do not necessarily signify that a scale composed of such items is unidimensional (Green, 1968). A better tool is an exploratory factor analysis. Though used mainly for data reduction, exploratory factor analysis also is useful in detecting dimensionality (Anderson, Basilevsky, & Hum, 1983). Therefore, to assess whether a single factor described the covariance structure adequately, a principal-factor analysis was performed on the total sample of students. Eigenvalues of the resulting covariance matrix of nine items were 5.727, .813, .688, .473, .386, .362, .292, .197, and .182. Only one eigenvalue exceeded 1, a criterion used conventionally to set the number of factors that can be extracted from a covariance matrix (Johnson & Wichern, 1982). Because the largest eigenvalue was 5.727, the proportion of variance explained by a single common factor solution is $5.727/9 = .636$. The second largest eigenvalue, .813, was a substantial drop in magnitude. Accordingly, the proportion of variance captured by a two-factor solution $[(5.727 +$

Table I. The Correlations and Means and Standard Deviations of the Nine FDR Scale Items

Variable	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
	Female (<i>n</i> = 341)								
Item 1	1.0000								
Item 2	.4980	1.0000							
Item 3	.4552	.7631	1.0000						
Item 4	.4625	.6401	.7221	1.0000					
Item 5	.5029	.3815	.4165	.4594	1.0000				
Item 6	.3340	.5454	.5778	.5775	.3830	1.0000			
Item 7	.5263	.4791	.5358	.5700	.5203	.4538	1.0000		
Item 8	.3120	.7564	.7325	.6167	.3459	.5803	.4736	1.0000	
Item 9	.3794	.4644	.4342	.4350	.3366	.4826	.4627	.4438	1.0000
Mean	4.7214	3.7859	3.7507	3.9091	4.6774	3.8889	4.5044	3.5308	4.4809
SD	.7832	1.2642	1.2647	1.2698	.8017	1.1854	.9413	1.2449	.9473
	Male (<i>N</i> = 237)								
Item 1	1.0000								
Item 2	.6938	1.0000							
Item 3	.6351	.8336	1.0000						
Item 4	.6505	.7202	.7270	1.0000					
Item 5	.6068	.6142	.5406	.6617	1.0000				
Item 6	.6063	.6730	.7042	.7074	.5638	1.0000			
Item 7	.6679	.6464	.6399	.7258	.6729	.6592	1.0000		
Item 8	.5525	.7605	.8098	.6925	.5638	.7566	.6422	1.0000	
Item 9	.5554	.6356	.5467	.5420	.5422	.5929	.6467	.5852	1.0000
Mean	4.3544	3.3924	3.2869	3.4473	4.2658	3.4979	4.0169	3.1392	4.0169
SD	1.0541	1.3443	1.3285	1.4241	1.0380	1.3007	1.2554	1.3845	1.2419

Table I. Continued

Variable	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
Item 1	1.0000								
Item 2	.6027	1.0000							
Item 3	.5558	.7989	1.0000						
Item 4	.5707	.6841	.7320	1.0000					
Item 5	.5784	.5079	.4937	.5749	1.0000				
Item 6	.4850	.6114	.6435	.6466	.4879	1.0000			
Item 7	.6215	.5712	.5982	.6590	.6222	.5684	1.0000		
Item 8	.4483	.7633	.7725	.6603	.4698	.6688	.5965	1.0000	
Item 9	.4975	.5586	.5045	.5050	.4728	.5497	.5832	.5270	1.0000
Mean	4.5709	3.6246	3.5606	3.7197	4.5087	3.7284	4.3045	3.3702	4.2907
SD	.9211	1.3109	1.3101	1.3532	.9276	1.2478	1.1065	1.3170	1.1008
Total (N = 578)									

Table II. Principal-Components Factor Analysis of the Nine-Item FDR Scale

Variable	One-factor solution		Two-factor solution		
	Estimated loading	Communality	Estimated loading	Communality	
Item 1	.741	.549	.741	.389	.701
Item 2	.857	.735	.857	-.241	.793
Item 3	.859	.739	.859	-.302	.830
Item 4	.847	.717	.847	-.050	.720
Item 5	.717	.514	.717	.463	.728
Item 6	.791	.626	.791	-.174	.656
Item 7	.805	.649	.805	.301	.739
Item 8	.828	.684	.828	-.387	.834
Item 9	.717	.514	.717	.157	.539
Eigenvalue	5.727		5.727	.813	
Cumulative proportion of variance explained	.636		.636	.727	

.813)/9 = .727] is a marginal improvement over variance explained by a one-factor model. The factor loadings and communalities of one- and two-factor solutions are presented in Table II. As expected, the communalities of the two-factor solution were higher than those of the single-factor solution, but the increase in the proportion of variance explained by the two-factor solution was only .091, which is modest, at best. Further inspection of transformed factor loadings from both orthogonal and oblique rotations for the two-factor model revealed no pattern suggesting that the two-factor model was preferred. Overall, therefore, results supported the conclusion that the FDR scale is unidimensional.

Selection of Measurement Model. Given the dimensionality opted for, a measurement model is needed for two reasons. First, given that the underlying construct is unobservable, investigation of the validities and reliabilities of multiple indicators or their linear composite requires a measurement model which establishes explicitly the linkage between the construct and its indicators. Second, the covariance structure of the indicators derived from the model allows statistical tests of the sample data (Greene & Carmines, 1980; Joreskog, 1971a). Random error measurement models that have been proposed are parallel, tau equivalent, and congeneric (Joreskog, 1969; Lord & Novick, 1968; see also Alwin & Jackson, 1979, for an overview). To find the most appropriate model, all three models were estimated, and their goodness-of-fit measures compared.

For a single-factor model with p observed variables, the variance–covariance matrix of the p observable variables can be expressed as functions of the latent factor, that is,

$$\Sigma = \Lambda\phi\Lambda' + \theta^2 \quad (1)$$

where Λ is a $p \times 1$ vector of factor loadings or pattern coefficients which relates the observed variables with the latent factor, Σ and Θ^2 are variance–covariance matrices of the observed variables and the disturbances, respectively, and ϕ is a scaler of the variance of the latent factor. The fitness of the model can be tested by comparing the covariance matrix derived from the model and that of the sample.

Without further restriction, the one-factor model is unidentified (Bollen, 1989; Joreskog, 1969). Using the congeneric model, the solution is to assign the metric of an observable variable to the latent variable by fixing one element of the pattern coefficient matrix to 1. If we assume further that the p observable variables measure the latent variables in the same metric, we specify each element of Λ as 1. The model then becomes tau equivalent. Further, if the variances of the disturbances are assumed to be the same for all observable variables, the model becomes parallel. To estimate the three measurement models—parallel, tau equivalent, and congeneric—a confirmatory factor analysis was performed in LISREL (Joreskog, 1969; Joreskog & Sorbom, 1988), using the sample data presented in Table I. Table III presents the resulting goodness-of-fit measures of three models, estimated by the method of maximum likelihood. As can be seen from Table III, none of the three models fits the data particularly well. The fit of the congeneric model, the least restrictive of the three, is relatively better than that of the other two. Even for the congeneric model, however, the chi-square is statistically significant, and none of the three relative fit measures—Tucker–Lewis coefficient, goodness-of-fit index (GFI), and adjusted goodness-of-fit index (AGFI)—is even close to .95, regarded as an indication of satisfactory fit (see Bollen, 1989).

There are two ways to remedy the situation. One is to postulate two or more factors, instead of one, which the multiple indicators represent. Another is to remove items from the analysis believed to be “noncongeneric” (Joreskog, 1971a). The first option was rejected for reasons noted previously. That is, exploratory factor analysis of the nine FDR items showed that only a small proportion of explained variance was gained by introducing the second factor into the model. In addition, a two-factor solution did not fit the data well either (GFI = .90). This could be due to the two factors being quite similar, since an oblique rotation for the two-

Table III. Goodness-of-Fit Measures of Confirmatory Factor Analysis of the Parallel, Tau-Equivalent, and Congeneric Models for Nine FDR Scale Items

	Parallel	Tau equivalent	Congeneric
χ^2	760.08	724.81	356.13
χ^2/df	17.68	20.71	13.19
Tucker-Lewis coef.	.83	.80	.88
GFI	.728	.725	.861
AGFI	.715	.647	.768

factor solution yielded a between-factors correlation of .681. As a result, the second option, deleting noncongeneric items, was adopted.

A Six-Item Congeneric Model. Results of the principal-factor analysis of the nine items on the FDR scale (see Table II) showed second factor loadings on Items 1 (spent a lot of money on her), 5 (she was drunk), and 7 (she had intercourse with other males). As shown in the Appendix, the remaining items (Nos. 2-4, 6, and 8-9) describe an interpersonal interaction, especially a sexual interaction, on a date. Even in the one item (No. 4) where a sexual interaction is not explicit, it is implicit. Substantively then, the six-item factor is interpretable as attitudes toward (i.e., acceptance-rejection of) the use of force to have sex in a sexual interaction with a date. On the other hand, among the three second-factor items, Item 1 describes an interaction between the pair in the vignette, as in Factor 1, but the interaction is impersonal, with no sexual interaction implied. Item 5 is neither an interpersonal nor a sexual interaction, and Item 7 describes sexual interactions, but not between the pair. It is unclear substantively what is measured by Items 1, 5, and 7—one reason for opting to delete them as noncongeneric items.

Given a six-item congeneric model, Table IV shows the estimates of variance-covariance structure for the three true score models. It is apparent from Table IV that the congeneric model fits the data better than the more restrictive parallel and tau-equivalent models. All the measures of goodness of fit improve dramatically from the more restrictive models to the congeneric model. The difference between the chi-square of the congeneric and that of the parallel model is $(282.81 - 70.28) = 212.53$, indicating a highly statistically significant improvement in the fit of the data. The GFI, AGFI, and Tucker-Lewis coefficient of the congeneric model are all in excess of .90, also suggesting a reasonable fit. Therefore, we conclude that the congeneric model provides the best fit to the data.

Table IV. Maximum-Likelihood Estimates of Parallel, Tau-Equivalent, and Congeneric Models of Six Indicators of Forcible Date Rape ($n = 576$)

Variable	Parallel		Tau equivalent		Congeneric	
	λ_i^a	θ_i^{2b}	λ_i^a	θ_i^2	λ_i	θ_i^2
Item 2	1	.578	1	.434	1 ^a	.400
Item 3	1	.578	1	.396	1.021	.341
Item 4	1	.578	1	.651	.934	.659
Item 6	1	.578	1	.658	.812	.687
Item 8	1	.578	1	.463	.989	.444
Item 9	1	.578	1	.838	.592	.749
Goodness-of-fit measure						
χ^2		282.81		212.34		70.28
χ^2/df		14.88		15.16		7.81
Tucker-Lewis coeff.		.91		.91		.96
GFI		.85		.88		.96
AGFI		.83		.82		.91

^aFixed parameters.

^bParameters are set to be equal.

Assessment of Reliabilities and Validities. An issue closely related to selection of a measurement model is assessment of reliability and validity. Both reliability and validity can be perceived as correlational linkage between an observable variable and a "true" variable that the observed variable is intended to measure (Alwin, 1991). Given the nature of latent variables being unobservable, reliability and validity can be assessed only if certain structural relationships between the observed and the unobserved are assumed (Greene & Carmines, 1979). Thus, assessment of reliability and validity must be based on a certain measurement model.

As noted previously, we viewed the six items as indicators of a theoretical construct, forcible date rape (or, more specifically, acceptance-rejection of use of force to have sex on a date). Therefore, these six measures were combined into a single composite. This is desirable, because the linear composite measure is more parsimonious, and because its reliability is always higher than any of its component variables, thus providing a more accurate representation of the latent construct. A composite variable is the weighted or unweighted linear combination of its component variables. When unweighted, and the measurement model is assumed to be parallel or tau equivalent, the reliability of the composite becomes the traditional Cronbach alpha coefficient (Greene & Carmines, 1979).

Table V. Estimated Reliabilities of Six FDR Scale Items and Unweighted Linear Composites

Variable	Parallel	Tau equivalent	Congeneric
Item 2	.645	.645	.767
Item 3	.645	.722	.801
Item 4	.645	.740	.640
Item 6	.645	.634	.559
Item 8	.645	.631	.744
Item 9	.645	.574	.382
Unweighted composite	.916	.916	.920

The estimated reliabilities of items and the unweighted linear composite based on the three models specified earlier are reported in Table V. In the parallel model, where the pattern coefficients are set to be 1 and error variances of all observed variables are equal, the reliabilities of all measures are equal, by definition. For the less restrictive tau-equivalent model, differences among the six items begin to emerge. For the congeneric model, which has been shown to be the most appropriate for current data, the estimated reliabilities differ across items. Items 2, 3, and 8 have higher reliabilities than the other three, implying that the former contribute more useful information than the latter about attitudes toward forcible date rape.

The reliability of an unweighted composite based on a parallel model is identical to Cronbach's alpha. If, instead, the model is congeneric or tau-equivalent, the Cronbach alpha provides a lower bound for the reliability of the unweighted composite based on these models (Bohrnstedt, 1983; Novich & Lewis, 1967). The results display such a pattern (see Table V). Having shown that the congeneric model provides a better fit to the data than the other two models, we concluded that .92 is the reasonable estimate of the reliability of the linear composite.

To be valid, a measure has to be reliable. But the reverse is not necessarily true. This relation is best expressed by the fact that the index of reliability places an upper bound for empirical validity (Alwin, 1991; Bohrnstedt, 1983). Since the overwhelming majority of reported rapes is by men, including college men, men would be expected to be less sensitive than women to the use of force to have sex. As one test of criterion validities of the six FDR items then, we tested the difference between the mean scores for males and those for females on these measures. As can be seen from Table I, the mean responses by females were higher than by males on all six items, and the differences were statistically significant (for

the six items, all t 's > 3.5). As expected, therefore, college females were more rejecting of the use of force to have sex than were males.

Two variables provide criteria for tests of validity. First, respondents were asked whether the behavior described in the vignette was rape, and they responded using a 5-point Likert scale ranging from "definitely" to "definitely not" rape (see Question 1 in the Appendix). If the six items are valid, the more uncertain people are that the male's behavior is rape (higher score), the more likely they would be expected to view the use of force as acceptable (i.e., score lower on FDR). On the other hand, the more definitely a person believes the male's behavior to be rape, the more likely they would view the use of force as unacceptable (i.e., score high on the FDR scale). Thus, strong negative relations between the response to the "Is it rape?" question and that to the six items would be expected. Pearson correlation coefficients between the variable and the six items and their unweighted linear components were $-.446$, $-.451$, $-.477$, $-.407$, $-.390$, $-.402$, and $-.511$. The complete consistency of negative correlations of items and their linear composite with the criterion variable is supportive. In addition, the magnitude of correlations is similar, which suggests that the factor is represented in the six items about equally and, thus, is supportive of the conclusion of unidimensionality of the FDR scale (see Bohrnstedt, 1969; Carmines & Zeller, 1983).

Second, right after respondents indicated how certain they were that the behavior described in the vignette was rape, they were asked "Where did the fault mainly lie?" (see Question 2 in the Appendix). The choices given were the male, the situation, society, and the female. Clearly, this variable measures the attitude toward culpability in a date rape situation. Though the choices are nominal, students not blaming the male as a group could be viewed as not blaming the male, the aggressor. If the six items and their composite are valid, we would expect persons not blaming the male to tend to consider the use of force by the male as more acceptable than those who blame the male. Again, negative correlations would be expected. Spearman correlation coefficients between the dichotomous variable of blame and the six items and their unweighted linear composite were $-.291$, $-.300$, $-.263$, $-.240$, $-.284$, $-.189$, and $-.309$. The uniformly negative correlations between this measure of blame and the six items and the composite also lend support for construct validity.

In summary, the greater rejection of the use of force to have sex by females than males and the consistently negative correlations between such judgment and judgments as to how certain they were that the rape described in the vignette really was rape and blaming the male, combined with high reliabilities, suggest the six items measure what they are supposed to, viz., attitudes toward forcible date rape.

The Weighting of the Linear Composite. Because a linear composite variable is constructed to represent its underlying construct, it is desirable to make it as reliable as possible. For an unweighted or equally weighted composite, each component variable contributes equally to the variation of the composite. Indeed, there is no need for unequal weighting when all the component variables have the same reliability, as is the case for the parallel model. However, when indicators have differential reliabilities, it is sometimes preferable to assign differential weights to component variables of a linear composite, to boost reliability (Alwin, 1973; Alwin & Jackson, 1979; Saris, de Pijper, & Mulder, 1978).

For a linear composite, $y = w'x$, where w' is a $p \times 1$ vector of weights and x is a $p \times 1$ vector of observed variables, the reliability of the composite can be expressed as

$$\rho_y^2 = \frac{w' \Lambda \phi \Lambda' w}{w' \Sigma w} \quad (2)$$

Substituting (1) into (2), we then have

$$\rho_y^2 = \frac{w' \Lambda \phi \Lambda' w}{w' \Lambda \phi \Lambda' w + w' \Theta^{-2} w} \quad (3)$$

It can be shown that when the weight w' are proportional to $\Lambda' \Theta^{-2}$, the product of pattern coefficients, and the inverse of error variance matrices, the reliability of the linear composite is maximal (Joreskog, 1971b; Maxwell, 1971). Because the elements of vector $\Lambda' \Theta^{-2}$ are uniform, the reliability of the unweighted composite of parallel measures is maximal. For tau-equivalent and congeneric measures, weighting proportional to $\Lambda' \Theta^{-2}$ produces the largest composite variable reliability.

Having established the congeneric model as the most suitable, a weighted linear composite was constructed. The reliability of this composite was .933. The reliability is computed by applying formula (3), where w' is assigned to be equal to $\Lambda' \Theta^{-2}$. Compared with the reliability of the unweighted composite based on the congeneric model (see Table V), the improvement is modest. As has been noted, when individual measures are well designed, the differences between weighted and unweighted linear composite variables is not very great (Alwin, 1973; Bohrnstedt, 1983). In the current case, the increase in reliability is moderate at best, and the unweighted composite has the merits of being both reliable and simple to construct.

Table VI. Maximum-Likelihood Estimates of the Congeneric Model of Six FDR Scale Indicators for Females and Males

Variable	Females ($n = 341$)		Males ($n = 237$)	
	λ_i	θ_i^2	λ_i	θ_i^2
Item 2	1 ^a	.414	1 ^a	.382
Item 3	1.025	.356	1.009	.314
Item 4	.909	.634	.967	.697
Item 6	.739	.758	.883	.581
Item 8	.962	.453	1.023	.426
Item 9	.468	.638	.694	.856
Goodness-of-fit measure				
χ^2	42.19		48.20	
χ^2/df	4.68		5.36	
Tucker-Lewis coeff.	.95		.94	
GFI	.96		.94	
AGFI	.91		.86	

^aFixed parameter.

Factorial Invariance Between Females and Males. Women students judged the use of force more unacceptable than did men on each of the six items. However, gender differences in mean responses could be present without the factor structure of the models for women and men necessarily being different. To test the latter, all three true score models were fit to the data, by gender. The equivalence of the measurement models for females and males was tested, following Joreskog (1971b). As was the case for the data as a whole, results of congeneric model fitting by gender (see Table VI) show that the congeneric model again provides the best fit for each gender. The patterns of magnitudes of estimated coefficients are remarkably similar for both genders (see Table VI). However, all goodness-of-fit measures show that the congeneric model fits less well for men than for women. As a result, the equivalence of the model for men and women must be rejected.

A formal statistical procedure to test the equivalence of certain properties of the measurement model across populations consists of a sequence of hypothesis testing, using chi-square (Joreskog, 1971b). Because chi-square increases with sample size, it has been noted that this procedure may be too sensitive to differences among sample variance-covariances when sample sizes are even moderately large (Alwin & Jackson, 1979). Nevertheless, for completeness, factorial invariance of congeneric models in the large female and male populations was tested (see Table VII). The first hypothesis

Table VII. Summary of Testing of Hypotheses of Similarity Between Measurement Models for Males and Females

Hypothesis	χ^2	df	<i>p</i> value
$\Sigma_f = \Sigma_m$	53.68	21	.000
$\Lambda_f = \Lambda_m$	105.63	23	.000
$\Lambda_f = \Lambda_{m'}$ $\Theta_f^2 = \Theta_m^2$	119.27	29	.000
$\Lambda_f = \Lambda_{m'}$ $\Theta_f^2 = \Theta_{m'}^2$			
$\phi_f = \phi_m$	124.02	30	.000
$(\Lambda/k)=1$	13.95	5	.016
$(\Lambda, \Theta^2/k)=1$	27.59	11	.004
$(\Lambda, \Theta^2, \phi/k)=1$	48.84	12	.001

in the procedure is that the covariance structures for males and females are equal. If this hypothesis were rejected, the next hypothesis is that there is only one factor in both populations, followed by the hypothesis of equivalence of factor loadings ($\Lambda_f = \Lambda_m$), and the equivalence of factor loading and error variances together ($\Lambda_f = \Lambda_m$ and $\Theta_f^2 = \Theta_m^2$). Finally, the equivalence of all model parameters ($\Lambda_f = \Lambda_m$, $\Theta_f^2 = \Theta_m^2$, $\phi_f = \phi_m$) is assessed. The last set of tests concerns the sameness of parameters given the factor number is determined. It is clear from Table VII that all these hypotheses can be rejected.

DISCUSSION

A measurement model was developed for nine items alleged to measure how acceptable-unacceptable is the use of force to have sex on a date. Based on an exploratory factor analysis (principal-components procedure), the magnitude of eigenvalues of the sample variance-covariance affirmed the likelihood of a single-factor solution. Based on this assumption, three true score measurement models — parallel, tau equivalent and congeneric — were fitted to the data. None of these models fit the variance-covariance of nine indicators very well. Based on substantive reasoning, three items were deleted. For the variance-covariance of the remaining six items, a congeneric model proved to be a reasonable fit. Thus, a six-item FDR scale consisting of Items 2, 3, 4, 6, 8, and 9 from the Appendix is recommended.

Unless the measurement model is incorporated explicitly into a structural model, which is the case for general structural equation models (Bollen, 1989; Joreskog, 1973), it is desirable to construct a linear composite variable as a proxy for its underlying construct. Whether used as a dependent variable or as a predictor, it is important that the composite variable be reliable and valid. For the unweighted composite of the six items, the measure of .92 suggests a reasonably high reliability. We also assigned differential weights to the individual items in the composite so that reliability was maximized. The maximal reliability of the weighted linear composite is .93. Given the complexity involved in constructing the weighted composite and the limited improvement in reliability, use of the unweighted composite is suggested.

Confirmatory factor analyses on female and male subsamples showed that the congeneric model fit both subsamples reasonably well. Moreover, the patterns of factor loadings and error variances were quite similar. However, a set of chi-square tests shows that the congeneric model was not the same for female and male subpopulations. This seeming contradiction may have occurred because of the sensitive nature of the chi-square test. In this situation, the chi-square statistic was used to test simultaneously the equality of two sets of parameters. Any difference between one pair of parameters will prompt rejection of equality of all parameters. Therefore, the test tends to be very conservative. Moreover, because chi-square tests the *exactness* of the model parameters across populations, when sample sizes are even moderately large, any minor difference will be detected. Thus, a large chi-square shows only that the hypothesis of exactly the same parameters must be rejected; it does not necessarily disprove their similarity. In other words, although there are differences between female and male populations, the overall pattern of their respective model measures is similar. One possible explanation of why they were found not to be the same is that there may be more variability in men's attitudes toward forcible date rape than in women's, and the 5-point Likert scale simply does not capture that difference fully. This possibility could be assessed by comparing 5-point responses with 7-point responses (e.g., adding "moderately acceptable" and "moderately unacceptable" responses).

Data on which the present paper was based did not permit addressing important issues such as the temporal stability and criterion validity of the FDR scale. In future research, test-retest data would be desirable, so as to assess temporal stability (Heise, 1969). Criterion validity might be addressed by assessing the discriminating/predictive power of FDR with respect to sexually coercive and noncoercive college men, i.e., those who self-report having lied or made false promises to have sex versus those who do not.

APPENDIX

A male and female college student go out on a date. Afterward, they go to his apartment and sit in front of the fireplace for a while and sip a glass of wine. He kisses her and, even though she resists his advances, uses his superior strength to force her to have sexual intercourse.

1. Would you call this rape?
 - A. Definitely
 - B. Probably
 - C. Not sure
 - D. Probably not
 - E. Definitely not

2. If rape did occur, where does the fault mainly lie?
 - A. The male
 - B. The situation—the apartment and alcohol were an avoidable risk and/or temptation
 - C. Society—for the way it socializes males and reinforces their dominant and assertive behavior
 - D. The female

For each of the conditions in 1–9 below, indicate how acceptable you consider the male's behavior in the above example.

1. If he had spent a lot of money on her.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - F. Definitely unacceptable

2. If she had gotten him sexually excited.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable

3. If she let him touch her breasts.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable

4. If they had dated each other for a long time.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable

APPENDIX Continued

5. If she was drunk.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable
6. If she was going to have intercourse with him and then changed her mind.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable
7. If she had intercourse with other males.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable
8. If she led him on.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable
9. If he was so sexually excited he couldn't stop.
 - A. Definitely acceptable
 - B. Mildly acceptable
 - C. Not sure
 - D. Mildly unacceptable
 - E. Definitely unacceptable

REFERENCES

- Alwin, D. F. (1973). The use of factor analysis in the construction of linear composites in social research. *Sociological Methods & Research*, 2, 191-214.
- Alwin, D. F. (1991). Research on survey quality. *Sociological Methods & Research*, 20, 3-29.
- Alwin, D. F., & Jackson, D. J. (1979). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology* (pp. 68-119). San Francisco: Jossey-Bass.
- Anderson, A. B., Basilevsky, A., & Hum, D. P. (1983). Measurement: Theory and techniques. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 231-287). San Diego, CA: Academic Press.
- Bohrnstedt, G. W. (1969). A quick method for determining the reliability and validity of multiple-item scales. *American Sociological Review*, 34, 542-548.
- Bohrnstedt, G. W. (1983). Measurement. In P. H. Rossi, J. D. Wright, & A. D. Anderson (Eds.), *Handbook of survey research* (pp. 70-121). San Diego, CA: Academic Press.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Burt, M. R. (1980). Cultural myths and support for rape. *Journal of Personality and Social Psychology*, 38, 217-230.

- Carmines, E. G., & Zeller, R. A. (1983). *Reliability and Validity Assessment*. Beverly Hills: Sage.
- Costin, F. (1985). Beliefs about rape and women's social roles. *Archives of Sexual Behavior, 14*, 319-325.
- Dull, R. T., & Giacomassi, D. J. (1987). Demographic correlates of sexual and dating attitudes: A study of date rape. *Criminal Justice and Behavior, 14*, 175-193.
- Feild, H. S. (1978). Attitudes toward rape: A comparative analysis of police, rapists, crisis counselors, and citizens. *Journal of Personality and Social Psychology, 36*, 156-179.
- Fischer, G. J. (1986). College student attitudes toward forcible date rape. I. Cognitive predictors. *Archives of Sexual Behavior, 15*, 457-466.
- Fischer, G. J. (1987). Hispanic and majority student attitudes toward forcible date rape as a function of differences in attitudes toward women. *Sex Roles, 17*, 93-101.
- Fischer, G. J. (1992). Sex attitudes and prior victimization predictors of college student sex offenders. *Annals of Sex Research, 5*, 53-60.
- Giarusso, R., Johnson, P., Goodchilds, J., & Zellman, G. (1979). *Adolescent cues and signals: Sex and assault*. Paper presented at meetings of the Western Psychological Association, San Diego, CA.
- Green, B. F. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology, Vol. 1*. Reading, MA: Addison-Wesley.
- Greene, V. L., & Carmines E. G. (1979). Assessing the reliability of linear composites. In K. F. Schuessler (Ed.), *Sociological methodology 1980* (pp. 160-175). San Francisco: Jossey-Bass.
- Johnson, R. A., & Wichern, D. W. (1982). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.
- Joreskog, K. G. (1971a). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.
- Joreskog, K. G. (1971b). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.
- Joreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*, pp. 85-112. New York: Academic Press.
- Joreskog, K. G., & Sorbom, D. (1988). LISREL VII. Scientific Software.
- Koralewski, M. S. & Conger, J. C. (1992). The assessment of social skills among sexually coercive college males. *Journal of Sex Research, 29*, 169-188.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Mahoney, E. R. (1983). *Human sexuality*. New York: McGraw-Hill.
- Maxwell, A. E. (1971). Estimating true scores and their reliabilities in the case of composite psychological tests. *British Journal of Mathematical and Statistical Psychology, 24*, 195-204.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1-13.
- Saris, W. E., de Pijper, M., & Mulder, M. (1978). Optimal procedures for estimation of factor scores. *Sociological Methods & Research, 7*, 85-106.