

Revisiting the Structural Validity of the Gender Role Conflict Scale

Bonnie Moradi, David M. Tokar, Michael Schaub, LaRae M. Jome,
and George S. Serna
University of Akron

Despite masculinity researchers' widespread use of the Gender Role Conflict Scale (GRCS; J. M. O'Neil, B. Helms, R. Gable, L. David, & L. Wrightsman, 1986), the structural validity of this instrument has recently been called into question. The authors revisited the status of the GRCS's structural validity via confirmatory factor analyses of both rationally and randomly developed item parcels as well as item-level data in a sample of 702 college men. Results indicated that, consistent with previous research, O'Neil et al.'s hypothesized oblique 4-factor model did not provide a conventionally good fit to the item-level data. However, as expected, superior (and conventionally good) fits to the data resulted when both rational and random parcel-level models were tested. Considered collectively, the results provide strong support for the structural validity of the GRCS and suggest that it is quite appropriate for masculinity researchers to score the GRCS for O'Neil et al.'s 4 factors.

Masculine gender role conflict is theorized to be rooted in men's gender role socialization, particularly men's adherence to the masculine mystique (i.e., a socially constructed set of values and beliefs that defines optimal masculinity) and their fear of femininity (O'Neil, 1981). More specifically, for men, adherence to rigid and sexist gender roles results in masculine gender role conflict because these restrictive roles prevent men from obtaining their full human potential and also restrict others from obtaining their full human potential (O'Neil, Good, & Holmes, 1995). Thus, masculine gender role conflict can be defined as "a psychological state in which socialized gender roles have negative consequences on the [man] or others" (O'Neil et al., 1995, p. 166).

The construct of gender role conflict has been of tremendous value in advancing theory and research on the psychology of men. A plethora of empirical research has linked gender role conflict with variables

that are critical to the psychology of men and masculinity. For example, gender role conflict has been shown to be related to men's overall psychological distress (Good et al., 1995), paranoia, interpersonal sensitivity, psychoticism, obsessive compulsivity (Good, Robertson, Fitzgerald, Stevens, & Bartels, 1996), anxiety (Courmoyer & Mahalik, 1995; Good & Mintz, 1990; Sharpe & Heppner, 1991), trait anger (Blazina & Watkins, 1996), intimacy-related variables (Fischer & Good, 1997; Good et al., 1995; Sharpe & Heppner, 1991; Sharpe, Heppner, & Dixon, 1995), relationship satisfaction (Sharpe et al., 1995), alexithymia (Fischer & Good, 1997), willingness to admit to alcohol use (Blazina & Watkins, 1996), attitudes toward seeking professional psychological help (Blazina & Watkins, 1996; Good, Dell, & Mintz, 1989), shyness (Bruch, Berko, & Haase, 1998), type of psychological defense use (Mahalik, Courmoyer, DeFranc, Cherry, & Napolitano, 1998), sense of entitlement (Hill & Fischer, 1998), sexual aggression against women (Rando, Rogers, & Brittan-Powell, 1998), vocational interests and career choice (Jome & Tokar, 1998; Tokar & Jome, 1998), and personality traits (Tokar, Fischer, Schaub, & Moradi, in press). Clearly, assessment of men's gender role conflict is important for researchers and practitioners interested in the psychology of men and masculinity.

Masculinity researchers have most often used the Gender Role Conflict Scale (GRCS; O'Neil, Helms, Gable, David, & Wrightsman, 1986) to operationalize and assess men's gender role conflict. The factor analytically derived GRCS consists of four subscales, each of which assesses one of four interrelated but distinct dimensions. The subscales or factors of the

Bonnie Moradi, David M. Tokar, Michael Schaub, LaRae M. Jome, and George S. Serna, Department of Psychology, University of Akron.

We would like to thank Ann R. Fischer for generosity in data lending and feedback on drafts of this article, Donald F. Walker for generosity in data lending, James R. Rogers for helpful comments on a draft of the article, Andrea F. Snell for statistical consultation, and Mike Plybon for creating the figure.

Correspondence concerning this article should be addressed to Bonnie Moradi or David M. Tokar, Department of Psychology, University of Akron, Akron, Ohio 44325-4301. Electronic mail may be sent to bmoradi@uakron.edu or dmt5@uakron.edu.

GRCS are (a) Success, Power, and Competition (SPC; assesses the extent to which one emphasizes personal achievement and control or authority over others); (b) Restrictive Emotionality (RE; assesses one's level of discomfort with emotional self-disclosure and with others' emotional expressiveness); (c) Restrictive Affectionate Behavior Between Men (RABBM; assesses one's level of discomfort with expressions of affection between men); and (d) Conflicts Between Work and Family Relations (CBWFR; assesses one's level of distress due to balancing the competing demands of work or school and family life).

Since its development, "the GRCS has proven itself to be a heuristic and fruitful contribution to the empirical literature on the male gender role" (Betz & Fitzgerald, 1993, p. 360). In light of the importance of the GRCS, some researchers have begun to assess the psychometric properties of the instrument. Betz and Fitzgerald recommended specifically the use of confirmatory factor analysis (CFA) to investigate the conceptual base of the gender role conflict model and the structural validity of the GRCS. CFA allows one to examine the fit of a specified model to a given data set (i.e., model-data fit) by comparing the sample covariance matrix of observed variables (computed from the input sample data) with the estimated population covariance matrix based on the proposed model. The degree of fit between the sample and estimated covariance matrices is summarized by the value of several fit indexes (Ullman, 1996).

Responding to this recommendation, Good et al. (1995) conducted several CFAs (including one-, two-, three-, four-, five-, and six-factor models) and found that, consistent with O'Neil et al.'s (1986) initial exploratory factor analysis results, the intercorrelated four-factor model provided the best fit to their data. On the basis of these results, Good et al. (1995) concluded that they had found "robust support for the factorial validity of the GRCS" (p. 6). In a cross validation and extension of this study, Rogers, Abbey-Hines, and Rando (1997) asserted that fit index values for the four-factor model obtained by Good et al. did not meet conventional criteria for acceptable model-data fit. Furthermore, Rogers et al. reported that the fit index values obtained from evaluation of the four-factor model in an independent sample also were below conventionally accepted criteria for a good fit. On the basis of their findings, Rogers et al. tempered Good et al.'s conclusions and suggested that their study provided tentative support for the structural validity of the GRCS. However, given that all fit index values calculated on Good et al.'s sample and Rogers et al.'s independent sample fell short of recommended cutoffs, Rogers et al.'s conclusion may have been premature as well.

An important methodological issue not addressed by Good et al. (1995) or Rogers et al. (1997) is that the value of fit indexes is influenced by indicator-per-factor (p/f) ratios (Ding, Velicer, & Harlow, 1995; Marsh, Hau, Balla, & Grayson, 1998; Williams & Holahan, 1994). More specifically, several studies have suggested that the value of goodness-of-fit indexes tends to decrease as the number of indicators per factor in a CFA increases (Anderson & Gerbing, 1984; Ding et al., 1995; Marsh et al., 1998; Williams & Holahan, 1994). The optimal number of indicators per factor remains debatable (Hall, Snell, & Foust, 1999); however, scholars have recommended using a minimum of three indicators per factor (e.g., Ding et al., 1995; Hall et al., 1999). This body of research suggests that previous studies' analyses, which used all 37 GRCS items as indicators and thus involved substantially larger p/f ratios (ranging from 6:1 [CBWFR subscale] to 13:1 [SPC subscale]), may have limited the chances of obtaining acceptable model-data fits for the GRCS. Thus, it seems that testing the structure of the GRCS as done in previous research is less than optimal.

Partly in response to problems posed by high p/f ratios, some researchers have used item parceling strategies (i.e., aggregating several items into one indicator) to reduce p/f ratios when conducting CFAs. Random parceling procedures are used to aggregate items within the same subscale or factor at random (with no attention to specific item content). According to rational parceling procedures, items within the same subscale or factor that are judged to share a secondary construct are aggregated into a new indicator (Hall et al., 1999).

Researchers typically construct parcels randomly; however, Hall et al. (1999) have recommended rational parceling procedures. In CFA, indicators of model fit and parameter estimates are based on covariance among indicators in the model. It is assumed that indicators assessing the same construct (i.e., latent factor) covary in part because of the shared influence of that construct. However, it is possible that, within a group of indicators that assess a particular factor or construct, certain indicators may also covary as a result of some secondary influence (i.e., a construct other than the factor the items are intended to assess, such as a method factor). Because parameter estimates and indicators of model fit are based on all shared variance, such secondary influence can contaminate model estimates. To prevent the occurrence of such model contamination, Hall et al. recommended rationally combining into the same parcel items that share a secondary construct. This procedure forces the influence of the secondary construct into the uniqueness (i.e., error) term of the

parcel and, in turn, models out the influence of the secondary construct from parameter estimates and indicators of model fit. Thus, Hall et al. suggested that rational parceling strategies have the advantage of mitigating possible model contamination resulting from secondary influences on indicators.

Rational parceling strategies may have some advantage over random parceling strategies; however, both strategies result in fewer indicators per factor. In turn, both strategies serve to reduce the potential limitations posed by large *p/f* ratios. Therefore, using these strategies in examining the structural validity of the GRCS seems warranted. Such an examination is especially important because high *p/f* ratios may have resulted in underestimates of the structural validity of the GRCS in previous research (i.e., Good et al., 1995; Rogers et al., 1997). Given the widespread use of the GRCS and its importance to the psychology of men and masculinity, reexamination of its structural validity using lower *p/f* ratios has the potential to contribute substantially to the emerging data on this popular instrument's utility.

In light of the methodological limitations discussed, the purpose of the current study was to reexamine the structural validity of the GRCS using both previously reported (i.e., Good et al., 1995; Rogers et al., 1997) procedures and procedures involving lower *p/f* ratios. That is, we examined the model-data fit for the GRCS with CFAs conducted at both the item and parcel levels. Because results may vary with rational and random parcel derivation, we tested models using both types of parcels. On the basis of the research reviewed here, we hypothesize that when the *p/f* ratio for the measurement model of the GRCS is reduced (i.e., when parcels rather than items are used as indicators), GRCS data will produce fit index values that meet the conventionally accepted criteria for a good fit. This information would allow researchers and practitioners to have greater confidence in the structural validity of the GRCS.

Method

Participants and Procedure

Participants consisted of three samples of men enrolled in undergraduate and graduate courses at a large public eastern-central university. These individuals completed the GRCS, along with other instruments, either individually or in classroom settings. Four different experimenters collected the data as part of three independent studies. Participants in the first sample ($n = 300$) completed the GRCS as part of a study examining the relations between masculinity-related

variables and the five-factor model of personality (Tokar et al., in press). The second sample ($n = 240$) completed the GRCS as part of an investigation of the relations among masculinity, vocational interests, and traditionality of career choice (Tokar & Jome, 1998). The third sample ($n = 162$) completed the GRCS as part of a study examining the dimensions underlying eight popular masculinity-related measures and their relations to sociosexuality (i.e., sexuality without commitment or emotional attachment; Walker, Tokar, & Fischer, 1999).

The three samples were combined to form the total sample ($N = 702$) for the current study. Participants ranged in age from 16 to 66 years ($M = 22.4$, $SD = 5.5$). In terms of ethnicity, most were White (83%); the remainder were African American (8%), Asian American (3%), multiracial (2%), Latino or Hispanic (1%), Native American or Indian (<1%), or of other ethnicities (2%). All participants were informed of procedures and provided written consent before taking part. On completion of the study, they received written debriefing materials and were given course credit for participation.

Instrument

The GRCS (O'Neil et al., 1986) measures personal dimensions of gender role conflict in men stemming from their fear of femininity (O'Neil, 1981). Respondents indicate their level of agreement with 37 items using a 6-point Likert-type scale (1 = *strongly disagree*, 6 = *strongly agree*); higher scores reflect more gender role conflict. A sample item is "I have difficulty telling others that I care about them." The GRCS is scored for the four factor analytically derived subscales described earlier: SPC (13 items), RE (10 items), RABBM (8 items), and CBWFR (6 items). O'Neil et al. reported 4-week test-retest reliabilities ranging from .72 to .86 and internal consistency reliabilities ranging from .75 to .85 for the four GRCS subscales. (Current internal consistency estimates for the subscales are presented in Table 1.) Construct validity is supported by significant relations in expected directions with depression (Courmoyer & Mahalik, 1995; Good & Mintz, 1990), help-seeking attitudes (Blazina & Watkins, 1996; Good et al., 1989), traditional male role norms, and psychological distress (Good et al., 1995).

Results

Means, standard deviations, internal consistency reliability estimates, and intercorrelations for the GRCS subscales and total scores are presented in Table 1. We conducted three CFAs to examine the model-data fit

Table 1
Means, Standard Deviations, Internal Consistency Reliabilities, and Intercorrelations for the Gender Role Conflict Scale Subscales and Total Score

Subscale or score	1	2	3	4	5	<i>M</i>	<i>SD</i>	α
1. Success, power, and competition	—	.41	.45	.38	.80	51.57	11.33	.86
2. Restrictive emotionality		—	.57	.30	.78	32.23	9.78	.87
3. Restrictive affectionate behavior between men			—	.30	.79	29.59	8.73	.86
4. Conflicts between work and family relations				—	.58	23.55	5.98	.76
5. Total score					—	136.9	27.02	.92

Note. All correlations are significant at $p < .001$.

for the GRCS using items, rationally derived parcels, and randomly derived parcels as indicators for the four factors of the GRCS. Data were analyzed with LISREL 8.14 (Jöreskog & Sörbom, 1993). For each model, covariance matrices were computed via PRELIS 2.14 and used as input. Various guidelines exist for estimating minimum sample sizes for structural equation modeling (of which CFA is a subset); however, one common recommendation is a minimum cases-to-parameter ratio of 5:1 (Bentler, 1993). Parameter estimates in the three models tested ranged from 36 to 80, suggesting that a sample size of 400 would be necessary. Our sample of 702 more than adequately met this recommendation.

The conventional chi-square statistic is easily distorted by sample size (see Ullman, 1996); therefore, values of several other fit indexes were examined, including the goodness-of-fit index (GFI), the adjusted GFI, the comparative fit index, the nonnormed fit index (also known as the Tucker-Lewis index), and the standardized root mean square residual. Values above .90 for the first four indexes, and values below .05 for the last index, indicate a good model-data fit (Ullman, 1996).

Model 1: Item-Level Model

To replicate Good et al.'s (1995) and Rogers et al.'s (1997) CFAs of the GRCS, we first tested a model us-

ing all 37 GRCS items as indicators. The four factors of the GRCS served as correlated latent variables, and the items corresponding to each factor or subscale were observed indicators. Specifically, SPC, RE, RABBM, and CBWFR had 13, 10, 8, and 6 indicators, respectively. Fit index values for this and all other models are presented in Table 2. Results indicated a model-data fit that, according to conventional criteria (e.g., $GFI \geq .90$), approached but would not quite be considered "good."

Model 2: Rationally Derived Parcel Model

A second model used rationally derived parcels as indicators for the four intercorrelated GRCS factors-subscales. For each factor-subscale, items judged by Bonnie Moradi and David M. Tokar to share a secondary construct were aggregated. Minor disagreements in item aggregation were discussed to mutual agreement, and the following parcels were created (GRCS item numbers and brief descriptors are indicated in parentheses): SPC1 (14 and 32; fear of failure), SPC2 (23, 24, and 28; competitiveness), SPC3 (21, 34, and 37; dominance), SPC4 (1, 5, 8, 12, and 18; success as a measure of personal worth), RE1 (13, 22, and 30; difficulty expressing feelings during sex), RE2 (2, 15, and 19; difficulty expressing feelings to others), RE3 (6 and 25; difficulty understanding feelings), RE4 (9 and 29; discomfort with/fear of sharing feelings

Table 2
Summary of Chi-Square and Fit Index Values for the Confirmatory Factor Analysis Models Tested

Model	χ^2	<i>df</i>	SRMR	GFI	AGFI	CFI	NNFI
Item level	2,042.87	623	.062	.86	.84	.85	.84
Rationally derived parcels	328.31	84	.049	.94	.91	.95	.93
Randomly derived parcels	281.54	84	.042	.95	.93	.96	.95

Note. $N = 702$ for all chi-squares. SRMR = standardized root mean square residual; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index; NNFI = nonnormed fit index.

with others), RABBM1 (16 and 35; discomfort with other men's affectionate behavior), RABBM2 (10 and 26; fear of expressing feelings to other men), RABBM3 (7 and 33; discomfort with expressing feelings to other men), RABBM4 (3 and 20; difficulty with expressing feelings to other men), CBWFR1 (11 and 27; work or school affects family life), CBWFR2 (4 and 17; work or school affects personal well-being), and CBWFR3 (31 and 36; work or school disrupts life in general).

As can be seen in Table 2, fit index values for this model surpassed the generally accepted cut-offs for the various indexes, demonstrating a very good model-data fit. In addition, each indicator loaded significantly ($p < .001$) and substantively on its intended factor (see Figure 1). Specifically, SPC loadings ranged from .65 to .84, RE loadings ranged from .69 to .84, RABBM loadings ranged from .70 to .85, and CBWFR loadings ranged from .60 to .85. Also revealed in Figure 1 are moderate to high GRCS factor intercorrelations (ranging from .37 to .67).

Model 3: Randomly Derived Parcel Model

A third model used randomly derived parcels as indicators of the four intercorrelated factors or subscales of the GRCS. For this model, subsets of items within each factor or subscale were randomly combined into parcels. To eliminate the possibility that any differences in fit between the rational and random parcel models were due to differences in p/f ratios, we created p/f ratios for the random parcel model to match those of the rational parcel model. Thus, four parcels were generated for SPC, RE, and RABBM, and three parcels were generated for CBWFR (item numbers are indicated in parentheses): SPC1 (1 and 12), SPC2 (8, 32, and 34), SPC3 (5, 21, and 37), SPC4 (14, 18, 23, 24, and 28); RE1 (13, 22, and 30), RE2 (9, 15, and 25), RE3 (6 and 19), RE4 (2 and 29); RABBM1 (16 and 26), RABBM2 (33 and 35), RABBM3 (3 and 10), RABBM4 (7 and 20); CBWFR1 (4 and 36), CBWFR2 (17 and 31), and CBWFR3 (11 and 27).

As can be seen in Table 2, fit index values for this model were very similar to those for Model 2. For both random and rational parcel models, fit index values suggested a very good model-data fit. Furthermore, as was revealed for Model 2, each indicator loaded significantly ($p < .001$) and substantively on its intended factor, and GRCS factor intercorrelations (ranging from .35 to .68) were moderate to high (see Figure 1). Specifically, SPC loadings ranged from .67 to .84, RE loadings ranged from .72 to .82, RABBM loadings ranged from .73 to .81, and CBWFR loadings ranged from .62 to .79.

Discussion

Despite the widespread use of the GRCS and its importance to the psychology of men and masculinity, the structural validity of this instrument has been called into question by recent research findings (Good et al., 1995; Rogers et al., 1997). More specifically, CFAs conducted by Good et al. and Rogers et al. yielded fit index values indicating a less than adequate fit to the data for the GRCS's conventional oblique four-factor model. However, our review of the measurement and research methodology literature suggested that Good et al.'s and Rogers et al.'s CFA results might have been underestimates of the GRCS's structural validity because these researchers performed their analyses using relatively high p/f ratios (ranging from 6–13 indicators per factor), which have been shown generally to result in poorer model-data fits (e.g., Ding et al., 1995; Marsh et al., 1998). In light of this methodological issue, we revisited the status of the structural validity of the GRCS by examining potential differences in the fit to the data of O'Neil et al.'s (1986) hypothesized oblique four-factor model using both rationally and randomly developed item parcels (to reduce p/f ratios) as well as item-level data, all within a new sample.

Consistent with Good et al. (1995) and Rogers et al. (1997), CFA results based on item-level data indicated that the oblique four-factor model did not provide a conventionally good fit to the data. Although the fit index values for the item-level analysis did not meet conventional cutoffs for a good fit, they did approach these cutoffs. For example, GFI, adjusted GFI, comparative fit index, and nonnormed fit index values, which ranged from .84 to .86, suggested a fit approaching that which would be considered "good" according to the conventional criterion of .90 or above. In light of the potential methodological limitation of using p/f ratios that may have resulted in underestimates of fit, the obtained fit index values for the current item-level data (as well as those reported by Rogers et al.) may be considered moderately supportive of O'Neil et al.'s model. In addition, values for fit indexes reported both herein and by Rogers et al. (i.e., adjusted GFI, nonnormed fit index, and standardized root mean square residual) were remarkably similar. Such consistent model-item data fits across independent samples attest to the GRCS's considerable structural stability in samples of college student men.

As expected, superior fits to the data resulted when parcel-level models were tested (i.e., when p/f ratios were reduced to three or four indicators per factor). In absolute terms, all fit index values (except the chi-square value, which, as indicated earlier, is notoriously sensitive to sample size) surpassed conventional good

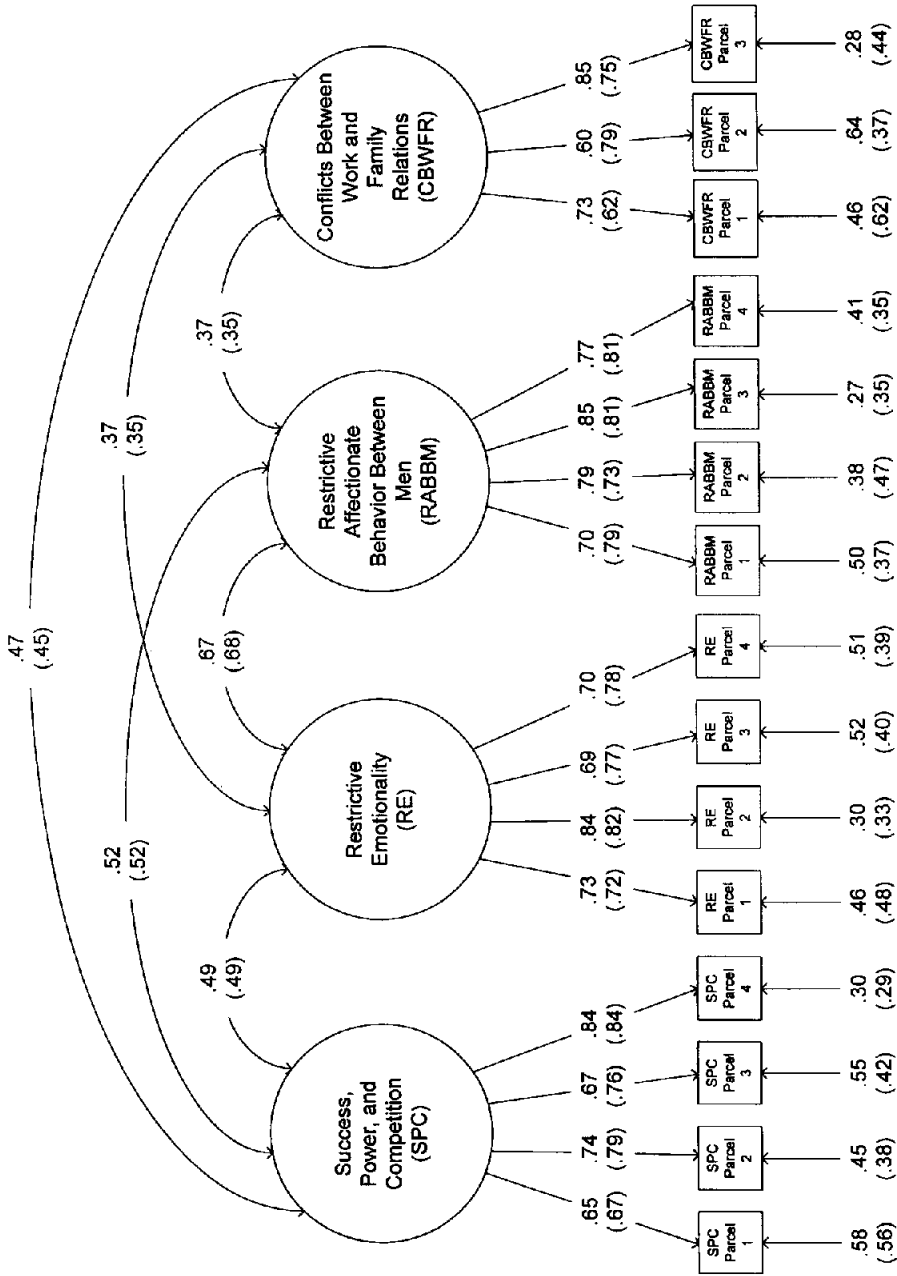


Figure 1. Model summary for rationally and randomly derived parcel models. Parentheses indicate values for the randomly derived parcel model. Values representing bidirectional paths between Gender Role Conflict Scale (GRCS) factors (e.g., SPC and RE) are factor intercorrelations. Values representing unidirectional paths from GRCS factors to corresponding parcel indicators are factor loadings. Values in the bottom two rows represent the error terms for the rationally and randomly derived parcels. All factor loadings and factor intercorrelations are significant at $p < .001$.

fit cutoffs for both the rationally and randomly derived parcel models. Furthermore, in both parcel-level models, each indicator loaded significantly and substantively (see Figure 1) on its intended GRCS factor. Considered collectively, results based on parcel-level data provide strong support for the structural validity of the GRCS and suggest that it is quite appropriate for masculinity researchers to score the GRCS for O'Neil et al.'s (1986) four factors. We encourage future masculinity researchers to consider using item parcels when conducting CFAs with a large number of indicators, particularly when p/f ratios are large and thus may jeopardize unnecessarily the likelihood of obtaining good model-data fits.

Of secondary interest to our study was the finding that rational and random parceling strategies produced little difference in fit index values. Hall et al. (1999) suggested that the effect of a secondary influence may contaminate the results of a CFA. Our results suggest that fit index values are not sensitive to such contamination. Thus, although rational parceling strategies may have resulted in a more accurate representation of the relationships between indicators and latent constructs (i.e., relationships uncontaminated by the influence of secondary constructs), they did not result in better fit index values with these data. Future masculinity researchers are encouraged to conduct additional tests comparing rational and random parcels of items from the GRCS and other masculinity-related measures.

In summary, despite potential limitations (e.g., a homogeneous and arguably nonrepresentative sample), our findings suggest empirical justification for scoring the GRCS as specified by the instrument's authors (O'Neil et al., 1986), at least in college student samples. On the basis of our findings, we recommend that future investigations of the GRCS's structural validity include analyses using item parcels. Also, we hope that the current results will allow future masculinity researchers to use the GRCS more confidently in research relating gender role conflict to additional variables of importance to the psychology of men and masculinity. Additional research (including investigations of the GRCS's structural validity) targeting underrepresented and underresearched populations of men (e.g., African Americans, Latinos, gay men, and bisexual men) is particularly needed.

References

- Anderson, J. C., & Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155-173.
- Bentler, P. M. (1993). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Betz, N. E., & Fitzgerald, L. F. (1993). Individuality and diversity: Theory and research in counseling psychology. *Annual Review of Psychology*, *44*, 343-381.
- Blazina, C., & Watkins, E. (1996). Masculine gender role conflict: Effects on college men's psychological well-being, chemical substance usage, and attitudes toward help-seeking. *Journal of Counseling Psychology*, *43*, 461-465.
- Bruch, M. A., Berko, E. H., & Haase, R. F. (1998). Shyness, masculine ideology, physical attractiveness, and emotional expressiveness: Testing a mediational model of men's interpersonal competence. *Journal of Counseling Psychology*, *45*, 84-97.
- Cournoyer, R. J., & Mahalik, J. R. (1995). Cross-sectional study of gender role conflict examining college-aged and middle-aged men. *Journal of Counseling Psychology*, *42*, 11-19.
- Ding, L., Velicer, W. F., & Harlow, L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, *2*, 119-144.
- Fischer, A. R., & Good, G. E. (1997). Men and psychotherapy: An investigation of alexithymia, intimacy, and masculine gender roles. *Psychotherapy*, *34*, 160-170.
- Good, G. E., Dell, D. M., & Mintz, L. B. (1989). Male role and gender role conflict: Relations to help seeking in men. *Journal of Counseling Psychology*, *36*, 295-300.
- Good, G. E., & Mintz, L. B. (1990). Gender role conflict and depression in college men: Evidence for compounded risk. *Journal of Counseling and Development*, *69*, 17-21.
- Good, G. E., Robertson, J. M., Fitzgerald, L. F., Stevens, M., & Bartels, K. M. (1996). The relation between masculine role conflict and psychological distress in male university counseling center clients. *Journal of Counseling and Development*, *75*, 44-49.
- Good, G. E., Robertson, J. M., O'Neil, J. M., Fitzgerald, L. F., Stevens, M., DeBord, K. A., & Bartels, K. M. (1995). Male gender role conflict: Psychometric issues and relations to psychological distress. *Journal of Counseling Psychology*, *42*, 3-10.
- Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*, 233-256.
- Hill, M. S., & Fischer, A. R. (1998, August). *Masculine entitlement, gender role traditionality, and sexual aggression*. Paper presented at the 106th Annual Convention of the American Psychological Association, San Francisco, CA.
- Jome, L. M., & Tokar, D. M. (1998). Dimensions of masculinity and major choice traditionality. *Journal of Vocational Behavior*, *52*, 120-134.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago: Scientific Software.
- Mahalik, J. R., Cournoyer, R. J., DeFranc, W., Cherry, M., & Napolitano, J. M. (1998). Men's gender role conflict

- and use of psychological defenses. *Journal of Counseling Psychology, 45*, 247–255.
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181–220.
- O'Neil, J. M. (1981). Patterns of gender role conflict and strain: Sexism and fear of femininity in men's lives. *Personnel and Guidance Journal, 60*, 203–210.
- O'Neil, J. M., Good, G. E., & Holmes, S. (1995). Fifteen years of theory and research on men's gender role conflict: New paradigms for empirical research. In R. F. Levant & W. S. Pollack (Eds.), *A new psychology of men* (pp. 164–206). New York: Basic Books.
- O'Neil, J. M., Helms, B., Gable, R., David, L., & Wrightsman, L. (1986). Gender Role Conflict Scale: College men's fear of femininity. *Sex Roles, 4*, 335–350.
- Rando, R. A., Rogers, J. R., & Brittan-Powell, C. S. (1998). Gender role conflict and college men's sexually aggressive attitudes and behavior. *Journal of Mental Health Counseling, 20*, 359–369.
- Rogers, J. R., Abbey-Hines, J. A., & Rando, R. A. (1997). Confirmatory factor analysis of the Gender Role Conflict Scale: A cross-validation of Good et al., 1995. *Measurement and Evaluation in Counseling and Development, 30*, 137–145.
- Sharpe, M. J., & Heppner, P. P. (1991). Gender role, gender-role conflict, and psychological well-being in men. *Journal of Counseling Psychology, 38*, 323–330.
- Sharpe, M. J., Heppner, P. P., & Dixon, W. A. (1995). Gender role conflict, instrumentality, expressiveness, and well-being in adult men. *Sex Roles, 33*, 1–18.
- Tokar, D. M., Fischer, A. R., Schaub, M., & Moradi, B. (in press). Masculine gender roles and counseling-related variables: Links with and mediation by personality. *Journal of Counseling Psychology*.
- Tokar, D. M., & Jome, L. M. (1998). Masculinity, vocational interests, and career choice traditionality: Evidence for a fully mediated model. *Journal of Counseling Psychology, 45*, 424–435.
- Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell, *Using multivariate statistics* (3rd ed., pp. 709–811). New York: HarperCollins.
- Walker, D. F., Tokar, D. M., & Fischer, A. R. (in press). What are eight popular masculinity-related instruments measuring? Underlying dimensions and their relations to sociosexuality. *Psychology of Men & Masculinity*.
- Williams, L. J., & Holahan, P. J. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work? *Structural Equation Modeling, 1*, 161–189.

Received July 15, 1999

Revision received October 25, 1999

Accepted November 5, 1999 ■