

Feminist Identity Development Measures: Comparing the Psychometrics of Three Instruments

Bonnie Moradi
Linda Mezydlo Subich
The University of Akron

Reliability and validity of three current instruments (Feminist Identity Scale [FIS], Feminist Identity Development Scale [FIDS], Feminist Identity Composite [FIC]) used to operationalize Downing and Roush's model of feminist identity development were compared. A sample of 245 women completed all three instruments, and a separate sample of 35 women repeated them over a 2-week interval. Only the FIC had acceptable internal consistency reliability for all subscales. Subscale stability for all instruments generally was moderate, except for Active Commitment. Subscale relations with perceived sexist events, self-esteem, social desirability, and preference for a male or female therapist generally were supportive of discriminant and convergent validity for all instruments. Content validity based on three judges' item evaluations suggested the FIDS fared best overall. Finally, confirmatory factor analysis procedures did not support definitively the structural validity of any of the instruments, but trends suggested the FIC, and perhaps the FIDS, were superior to the FIS.

Downing and Roush's (1985) model of feminist identity development has resulted in important theoretical and empirical work in the psychology of women (Moradi, Subich, & Phillips, 2002 [this issue]). Thus, assessing feminist identity development is important for researchers and practitioners. Downing and Roush's model is operationalized by the revised Feminist Identity Scale (FIS-R) (Rickard, 1989) and the Feminist Identity Development Scale (FIDS) (Bargad & Hyde, 1991). Because of psychometric concerns about each of these measures, Fischer et al. (2000) created the Feminist Identity Composite (FIC) from selected FIS and FIDS items.

Rickard (1989) developed and revised the FIS to assess cognitive and affective components of attitudes that a woman may have about herself as a feminist. Active Commitment, which Rickard perceived to be a behavioral manifestation of Synthesis, is not assessed by the revised FIS. The revised FIS has been used in several studies (e.g., Fischer et al., 2000; Juntunen,

The authors thank Ann R. Fischer for feedback on an earlier version of this manuscript and Andrea F. Snell for statistical consultation. Correspondence concerning this research should be addressed to Bonnie Moradi, Department of Psychology, P.O. Box 112250, University of Florida, Gainesville, FL 32611-2250; phone: 352-392-0601; fax: 352-392-7985; moradib@ufl.edu; or Linda Mezydlo Subich, Department of Psychology, The University of Akron, Akron, OH 44325-4301; phone: 330-972-8379; fax: 330-972-5147; e-mail: subich@uakron.edu.

THE COUNSELING PSYCHOLOGIST, Vol. 30 No. 1, January 2002 66-86
© 2002 by the Division of Counseling Psychology.

Atkinson, Reyes, & Gutierrez, 1994), but psychometric data on it are limited. Fischer et al. (2000) reported that their four-factor solution for the revised FIS accounted for only 29% of the total item variance. Further, internal consistency reliability estimates have been found repeatedly to be problematic; they typically are lowest for Passive Acceptance (.54 [Fischer et al., 2000] to .69 [Juntunen et al., 1994]) and highest for Revelation (.78 [Fischer et al., 2000] to .85 [Juntunen et al., 1994]). Finally, little evidence of discriminant or convergent validity is available for the revised FIS. Fischer et al. reported that revised FIS Passive Acceptance and Revelation subscale scores correlated most highly with their corresponding subscales of the FIDS, but revised FIS Embeddedness-Emanation and Synthesis subscales correlated equally or more strongly with FIDS Active Commitment than with their corresponding FIDS subscales. Due to its primacy in the research literature, hereafter the revised FIS shall be referred to as the FIS.

The FIDS (Bargad & Hyde, 1991) consists of self-descriptive statements that reflect all five stages of Downing and Roush's (1985) model. As with the FIS, concerns about the structural validity of the FIDS have been raised. Bargad and Hyde, as well as Gerstmann and Kramer (1997), reported some empirical support for a five-factor structure, but Bargad and Hyde suggested that future analyses could result in collapsing the Synthesis and Active Commitment subscales, and Gerstmann and Kramer, who specified a five-factor structure a priori, found that a five-factor solution accounted for only 40% of the variance in the FIDS. Additional support for collapsing the Synthesis and Active Commitment subscales has been reported by Cash, Ancis, and Strachan (1997). Yet, support for a four-factor structure that includes Passive Acceptance, Embeddedness-Emanation, Synthesis, and Active Commitment has been reported (Ng, Dunne, & Cataldo, 1995), and Fischer et al. (2000) reported that a three-factor solution was the most interpretable for the FIDS but accounted for only 27% of the total item variance. As with the FIS, scale reliability is an issue with the FIDS. The Synthesis subscale has been reported repeatedly to have unacceptable internal consistency reliability (.48 [Fischer et al., 2000] to .65 [Bargad & Hyde, 1991]). Highest internal consistencies typically have been reported for the Passive Acceptance (.74 [Fischer et al., 2000] to .85 [Bargad & Hyde, 1991]) and Active Commitment (.78 to .81 [Gerstmann & Kramer, 1997]) subscales. Finally, support for the validity of the FIDS exists. Correlations between FIDS scales and social desirability have been nonsignificant or negligible (Bargad & Hyde, 1991; Gerstmann & Kramer, 1997). In general, FIDS subscales correlated most highly with corresponding FIS subscales (Fischer et al., 2000). These findings are not entirely consistent, as the FIDS Embeddedness-Emanation correlated as strongly with the FIS Revelation as with its corresponding FIS subscale (Fischer et al., 2000).

Due to psychometric concerns about the FIS and the FIDS, Fischer et al. (2000) developed the FIC. The FIC is a composite of 20 FIS and 13 FIDS items and was created to capitalize on the strengths and eliminate some of the weaknesses of these two instruments. Initial psychometric data for the FIC are promising. Fischer et al. reported that the FIC adhered to a five-factor structure that resembled Downing and Roush's (1985) model and accounted for 36% of the total item variance. Fit index values from a confirmatory factor analysis suggested an excellent model-data fit. Internal consistency estimates for the subscales of the composite measure ranged from .68 for Synthesis to .84 for Embeddedness-Emanation in the first sample and from .71 for Synthesis to .86 for Embeddedness-Emanation in the second sample. Fischer et al. also reported that subscales of this measure were negligibly correlated with social desirability but related, as expected, to ego identity status, perceptions of sexist events, and involvement in women's organizations. Yet, no further research on this composite measure has been conducted.

Given the limited and mixed psychometric data on these three instruments, research on feminist identity development would be advanced if the psychometric properties of these measures were examined and compared. Using a single sample to address the possibility that previously reported differences in the psychometric properties of these measures may have been a function of sample-specific variables, analyzing a broad array of psychometric properties across instruments and using confirmatory factor analytic (rather than exploratory) methods may provide a clearer picture of the strengths and weaknesses of each measure.

Thus, this research examines with a single sample each of the three available operationalizations of Downing and Roush's (1985) model for evidence of (a) internal consistency and test-retest reliability, (b) discriminant validity, (c) convergent validity, (d) content validity, and (e) structural validity. More specifically, subscales for each instrument are examined for acceptable (i.e., $\geq .70$) internal consistency estimates (Nunnally, 1978). Next, the correlations between feminist identity development subscales and social desirability are examined for evidence of discriminant validity, and those between feminist identity development subscale scores and perceived sexist events, self-esteem, and preference for a male or female therapist are examined for evidence of convergent validity. Specifically, Downing and Roush suggested that Passive Acceptance is related to a denial of sexism, whereas later stages of feminist identity development, particularly Revelation, are related to the recognition of sexism in women's lives. Similarly, Passive Acceptance may be related negatively, whereas other feminist identity development scores may be related positively, to the amount of perceived stress associated with sexism. Downing and Roush also stated that the transition from Passive Acceptance to the later stages of feminist identity development is related to

higher levels of self-esteem, and McNamara and Rickard (1989) suggested that Synthesis and Active Commitment scores should be related to higher levels of self-esteem. Further, consistent with Downing and Roush's conceptualization that Revelation and Embeddedness-Emanation are marked by negative attitudes and feelings toward men and positive attitudes and feelings toward women, respectively, McNamara and Rickard suggested that high Passive Acceptance scores should be related to preference for a male therapist, whereas high Revelation or Embeddedness-Emanation scores should be related to a preference for a female therapist, and Synthesis and Active Commitment scores should be related to no preference for therapist gender. In addition, the interrelations among FIS and FIDS subscales are examined for evidence of discriminant and convergent validity; relations of FIC subscales to FIS and FIDS subscales reflect item overlap and are not interpreted. To assess content validity, judges' assignments of items to their intended subscales are examined, and confirmatory factor analysis is used to examine the structural validity of each instrument. Finally, short-term test-retest reliability is examined with a separate sample.

METHOD

Participants

To extend the generalizability of these findings, a sample of 159 undergraduate and 86 faculty and staff women ($N = 245$) from a large midwestern university were recruited. Undergraduate women were enrolled in psychology courses and received extra credit for participating. We selected randomly 295 faculty and staff women from the university directory, invited them to participate, and informed them that if they chose to participate, they would be entered into a lottery for five \$50 cash awards; awards were distributed subsequent to data collection. To protect participants' anonymity and confidentiality and improve participation rates, faculty/staff women were not asked to identify themselves as faculty or staff members. Three reminder postcards were mailed out, approximately 2 to 3 weeks apart, to all faculty/staff women. Eighty-six of the faculty/staff women completed surveys, resulting in a 29% return rate. Data from a portion of this sample ($N = 191$) were included in Moradi and Subich (2002 [this issue]).

Participants ranged in age from 16 to 67 years ($M = 30.24$, $SD = 12.68$). Seventy-nine percent of the sample identified as White, 13% African American, 4% Asian American, 2% Latina, 2% multiracial or other racial/ethnic group, and less than 1% Native American. Fifty-seven percent of participants were single and 42% were married or in a committed relationship. Forty-

seven percent of the participants identified as middle class, 31% working class, 17% upper-middle class, 4% lower class, and less than 1% upper class. Approximately 69% of the sample had a high school degree, 11% had a bachelor's/associate degree, 14% had a master's degree, and 6% had a doctorate. Overall, this sample's mean scale scores on the instruments used in this study approximated corresponding scale means obtained in previous studies (details available upon request). Stability data were gathered from a separate sample of 35 undergraduate women drawn from the same courses and institution as the primary sample. These women completed the FIS and FIDS twice over an interval of approximately 2 weeks. Their mean age was 22.37 years ($SD = 6.62$), and most were in their 1st year of school (49%) and White (65%).

Instruments

FIS. The revised FIS is described by Juntunen et al. (1994) as a 37-item Likert-type scale (1 = *strongly disagree* to 5 = *strongly agree*) with subscales corresponding to the first four stages of Downing and Roush's (1985) feminist identity development model (i.e., Passive Acceptance [PA], Revelation [R], Embeddedness-Emanation [EE], and Synthesis [S]). Subscale items' ratings are averaged to obtain a subscale score; higher scores indicate greater agreement with the corresponding feminist identity stage. Alpha coefficients for PA, R, EE, and S were .60, .78, .65, and .69, respectively (Juntunen et al., 1994). Three-week test-retest reliabilities for PA, R, EE, and S on the original FIS ranged from .83 to .93, and alphas exceeded .85 (Rickard, 1989).

FIDS. The FIDS is a 39-item Likert-type scale (1 = *strongly disagree* to 5 = *strongly agree*) with subscales corresponding to the five stages of Downing and Roush's (1985) model. Subscale items' ratings are averaged to obtain a subscale score; higher scores indicate greater endorsement of the corresponding feminist identity stage. Gerstmann and Kramer (1997) reported alphas of .74, .63, .71, .51, and .78 at Time 1 (beginning of semester), .80, .70, .83, .55, and .81 at Time 2 (end of semester), and 3-month test-retest reliabilities of .85, .75, .82, .69, and .85 for PA, R, EE, S, and AC, respectively.

FIC. Fischer et al. (2000) combined 33 items, 20 from the FIS and 13 from the FIDS, to create the FIC. We obtained subscale scores by averaging item ratings for each subscale. Higher mean scores for each subscale indicate greater agreement with the corresponding feminist identity stage. Fischer et al.

reported alphas of .75, .80, .84, .68, and .77 for PA, R, EE, S, and AC, respectively. No stability data are available for the FIC.

Balanced Inventory of Desirable Responding–Version 6. The BIDR-6 includes two subscales, Self-Deceptive Enhancement (SDE), which reflects “the tendency to give honest but inflated self-descriptions” (Paulhus, 1994, p. 2), and Impression Management (IM), which reflects “the tendency to give inflated self-descriptions to an audience” (Paulhus, 1994, p. 2). Each subscale consists of 20 Likert-type items (1 = *not true* to 7 = *very true*). One point was awarded for each 6 or 7 response (Paulhus, 1994). Thus, subscale scores had a possible range of 0 to 20, with higher scores indicating greater socially desirable responding. A two-factor structure corresponding to SDE and IM was supported (Paulhus, 1994). Internal consistencies ranged from .65 to .75 for SDE and from .75 to .80 for IM (Paulhus, 1994). Five-week test-retest reliability was .69 for SDE and .77 for IM. SDE related positively to optimism and positive reappraisal. IM related positively to lie scales and role-playing measures (Paulhus, 1994). Our internal consistency estimates for SDE and IM approximated values reported by Paulhus (i.e., .63 and .70, respectively).

Schedule of Sexist Events. The SSE (Klonoff & Landrine, 1995) consists of 20 items that assess the perceived frequency and appraisal of sexist discrimination. Using a Likert-type scale (1 = *the event never happened* to 6 = *the event happened almost all of the time*), participants rate items for the frequency of sexist events within one’s lifetime (SSE Lifetime) and within the past year (SSE Recent). In addition, using a Likert-type scale (1 = *not at all stressful* to 6 = *extremely stressful*), participants rate the stressfulness of the events reflected in SSE items (SSE Appraisal). Ratings across items are added to obtain scores for each of the three dimensions of the SSE; higher scores indicate greater frequency of perceived sexist events. Internal consistency estimates for the SSE Lifetime, SSE Recent, and SSE Appraisal dimensions were .92, .90, and .93, respectively (Landrine & Klonoff, 1997). Test-retest coefficients are available only for the Lifetime ($r = .70$) and Recent ($r = .63$) dimensions (Klonoff & Landrine, 1995). The SSE related positively to psychological distress (Landrine & Klonoff, 1997) and related negligibly or nonsignificantly to social desirability (Fischer et al., 2000). In this study, alphas for SSE Lifetime, SSE Recent, and SSE Appraisal were .92, .91, and .92, respectively.

Rosenberg Self-Esteem (RSE). We used the 4-point Likert-type (1 = *strongly disagree* to 4 = *strongly agree*) version of the RSE (Rosenberg, 1965; Wylie, 1989). Five of the 10 items are reverse scored and ratings are summed

across items. Scores range from 10 to 40; lower scores indicate lower self-esteem (Wylie, 1989). Alphas have ranged from .74 to .87 and test-retest reliabilities from .63 to .91 (Wylie, 1989). The RSE has been linked negatively to depressive affect, anxiety, psychosomatic symptoms, and interpersonal insecurity (Wylie, 1989). The RSE alpha in this study was .87.

Therapist preference. All participants were asked, "If you were to choose a therapist, would you prefer to have a female or a male therapist, or would you have no preference?"

Procedures

Procedures were described in writing to participants and written consent was obtained. To limit the extent to which responding to feminist identity development items influenced participants' responses to items assessing variables used in the convergent and discriminant validity analyses, we used the following order of instruments in the survey packet: SSE, RSE, therapist preference, BIDR, FIDS, and FIS. Due to missing data, five cases were eliminated, resulting in a total sample size of 240. Because five women did not respond to the therapist preference question, analyses involving preference for therapist had an *N* of 235.

To assess content validity, Fischer et al.'s (2000) procedures were used. That is, three trained judges assigned blindly and separately the items of the FIS, FIDS, and FIC to the appropriate feminist identity stage as outlined by Downing and Roush (1985). Judges were three female doctoral students in counseling psychology. None reported prior familiarity with the FIDS, FIS, or FIC. Each judge read Downing and Roush's article and was given two tables, one from Rickard (1990) and one from Bargad and Hyde (1991). The first author discussed the model with the judges and defined the characteristics of each stage. Then, for each instrument, each judge sorted items typed on index cards into appropriate subscales. Each judge's total item hit rates were calculated for each instrument.

RESULTS

Internal consistency reliability. Internal consistency estimates and corrected item-total correlations for the FIS, FIDS, and FIC are reported in Table 1. All internal consistencies for the FIC were above the generally accepted cutoff of .70 (Nunnally, 1978), but the FIS's PA and EE and the FIDS's R and S yielded alphas below this cutoff. Further, 2 of 33 (6%) FIC items, 6 of 39

TABLE 1: Internal Consistency Estimates and Item-Total Correlations for the FIS, FIDS, and FIC

Subscale	α			Corrected Item-Total Correlation Range (mdn)		
	FIS	FIDS	FIC	FIS	FIDS	FIC
Passive acceptance	.67	.79	.74	.24-.53 (.46)	.24-.53 (.44)	.36-.56 (.45)
Revelation	.80	.64	.76	.19-.66 (.40)	.19-.46 (.42)	.27-.64 (.46)
Embeddedness-emanation	.69	.76	.84	.05-.64 (.53)	.37-.57 (.48)	.56-.73 (.68)
Synthesis	.77	.52	.73	.18-.54 (.46)	.21-.39 (.27)	.29-.61 (.52)
Active commitment	—	.77	.77	—	.35-.61 (.46)	.35-.58 (.42)

NOTE: $N = 240$. FIS = Feminist Identity Scale; FIDS = Feminist Identity Development Scale; FIC = Feminist Identity Composite.

(15%) FIDS items, and 7 of 37 (19%) FIS items correlated less than .30 with their intended subscales.

Test-retest reliability. Two-week test-retest reliabilities for the FIS subscales were moderate to good ($r_{PA} = .64$, $r_R = .74$, $r_{EE} = .86$, $r_S = .64$). For the FIDS, comparable stability was noted for PA ($r = .77$), R ($r = .77$), and EE ($r = .79$), but S ($r = .50$) and Active Commitment (AC) ($r = .38$) were problematic. Stability for FIC AC ($r = .36$) was problematic ($r_{PA} = .65$, $r_R = .71$, $r_{EE} = .80$, $r_S = .70$).

Discriminant and convergent validity. Due to the number of correlations, alpha was adjusted to .01. Correlations between BIDR SDE and IM, and subscales of the FIS, FIDS, and FIC, were nonsignificant (see Table 2).

All FIDS subscales correlated highest with corresponding FIS subscales. Most FIS subscales correlated highest with corresponding FIDS subscales. However, FIS S correlated more highly, but in the expected directions, with FIDS PA and AC than with FIDS S.

Correlations among the subscales within each instrument suggested some problems in assessing a linear model of feminist identity development. Such a model would suggest higher correlations among measures of adjacent stages than among measures of nonadjacent stages. This was not always the case. For example, the correlations between FIS PA and FIS R and between FIDS EE and FIDS S were nonsignificant, and the correlation between FIC EE and FIC AC was somewhat greater than that between FIC S and FIC AC.

Nine of 12 (75%) correlations between the FIS subscales and dimensions of the SSE were significant. For both the FIDS and FIC, 11 of 15 (73%) correlations with SSE dimensions were significant. All significant correlations

TABLE 2: Intercorrelations and Descriptive Statistics for Major Variables

Variables	1a	1b	1c	1d	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	4	5a	5b	5c	6	M	SD	
1. Feminist Identity Scale																						
a. Passive acceptance																					3.13	.80
b. Revelation	-.03																				2.87	.59
c. Embeddedness-emanation	-.18*	.43**																			3.05	.69
d. Synthesis	-.29**	.18*	.38**																		3.87	.51
2. Feminist Identity Development Scale																						
a. Passive acceptance	.53**	-.02	-.22**	-.43**																	2.43	.61
b. Revelation	-.05	.63**	.38**	.17	-.06																3.06	.62
c. Embeddedness-emanation	-.11	.48**	.50**	.20*	-.10	.56**															2.79	.63
d. Synthesis	-.12	-.16	.04	.39**	-.23**	-.13	-.06														3.91	.54
e. Active commitment	-.27**	.32**	.46**	.44**	-.44**	.43**	.56**	.12													3.17	.56
3. Feminist Identity Composite																						
a. Passive acceptance	.61**	.02	-.17	-.37**	.89**	-.01	-.04	-.23**	-.34**												2.69	.72
b. Revelation	.02	.93**	.32**	.08	.03	.60**	.45**	-.18*	.27**	.05											2.66	.68
c. Embeddedness-emanation	-.27**	.28**	.90**	.42**	-.34**	.27**	.48**	.15	.51**	-.27**	.20*										3.03	.90
d. Synthesis	-.24**	.03	.23**	.86**	-.41**	.01	.00	.37**	.26**	-.36**	-.06	.25**									4.24	.61
e. Active commitment	-.30**	.29**	.46**	.61**	-.45**	.40**	.51**	.18*	.94**	-.35**	.22*	.52**	.40**								3.40	.54
4. Rosenberg Self-Esteem Scale	-.08	-.08	.07	.28**	-.19*	-.04	.01	.10	.18*	-.11	-.11	.07	.28**	.19*							33.80	5.91
5. Schedule of Sexist Events																						
a. Recent	-.03	.37**	.25**	.13	-.09	.33**	.19*	-.09	.21*	-.07	.38**	.19*	.05	.22*	-.12						38.54	14.38
b. Lifetime	-.16	.39**	.36**	.26**	-.17*	.37**	.25**	.03	.30**	-.16	.36**	.34**	.15	.30**	-.08	.81**					48.94	16.22
c. Appraisal	-.19*	.37**	.35**	.29**	-.25**	.41**	.28**	-.01	.31**	-.20*	.33**	.35**	.18*	.33**	-.04	.67**	.76**				46.49	18.94
6. Balanced Inventory of Desirable Responding																						
a. Self-deceptive enhancement	.00	-.10	.00	.02	-.11	-.09	-.09	.01	.03	-.05	-.10	.01	.09	.01	.29**	-.06	-.11	-.13			4.17	2.63
b. Impression management	-.01	-.08	-.07	.00	-.01	-.12	-.07	-.01	-.03	.06	-.06	-.08	.04	-.01	.21*	-.16	-.21*	.11	.40**		4.25	2.79

NOTE: $N = 240$. Decimal points have been eliminated from all correlation coefficients in this table.* $p < .01$. ** $p < .001$.

were in the expected direction (i.e., negative for PA, positive for all other subscales). For the FIS, correlations of SSE dimensions were small or nonsignificant with PA, were highest with R and EE, and dropped with S. For the FIDS, relations of SSE dimensions were small or nonsignificant with PA, were highest with R, dropped slightly with EE and AC, and were nonsignificant with S. For the FIC, relations of SSE dimensions were small or nonsignificant with PA, high with R, EE, and AC, and small or nonsignificant with S.

One of four (25%) correlations between the FIS subscales and self-esteem was significant; FIS S was related positively to self-esteem. Two of five (40%) correlations between the FIDS subscales and self-esteem were significant; FIDS PA was related negatively and FIDS AC was related positively to self-esteem. Finally, two of five (40%) correlations between the FIC subscales and self-esteem were significant; FIC S and AC were related positively to self-esteem. Thus, although each of the instruments yielded some relations to self-esteem that corresponded to expectations, the majority of the examined relationships were nonsignificant.

As seen in Table 3, only 4% of participants (i.e., 9 women) reported that if they were to choose a therapist, they would prefer to have a male therapist. Thus, significance tests were conducted only between women who reported a preference for a female therapist ($n = 92$; 41%) and those who reported having no therapist preference ($n = 134$; 56%). To examine differences in the set of feminist identity development scores between the two groups, three 2-way ANOVAs (one for each instrument) with one between (i.e., therapist preference) and one within (i.e., feminist identity development subscales) factor were conducted. The tests of the interaction effects in these analyses indicated a significant difference in feminist identity development scores by group for the FIS, $F(2, 473) = 4.69$; $p < .01$; $\eta^2 = .021$, and FIDS, $F(4, 689) = 5.78$; $p < .01$; $\eta^2 = .025$, but not for the FIC. Follow-up one-way ANOVAs (with alpha adjusted to .01) for the FIS and FIDS indicated that women who preferred a female therapist scored significantly higher on FIDS R, $F(1, 224) = 11.37$; $p < .01$; Cohen's $d = .46$, FIDS EE, $F(1, 224) = 11.65$; $p < .01$; Cohen's $d = .47$, and FIS EE, $F(1, 224) = 7.00$; $p < .01$; Cohen's $d = .35$, than did women who had no therapist preference. Table 3 presents means and standard deviations for feminist identity development scores by therapist preference (ANOVA findings are available from the authors).

Content validity. For the FIS, hit rates ranged from 73% to 87% ($Mdn = 78%$; $M = 79%$). Across the three judges, items from the R subscale were misplaced 31% of the time (13 of 42 times; placed in S 4 times, in EE 9 times), followed by S items that were misplaced 22% of the time (8 of 36 times; placed in EE 7 times, in PA once) and EE items that were misplaced 11% of

TABLE 3: Descriptive Data on Feminist Identity Development Subscale Scores and Preference for a Male or Female Therapist

<i>Instrument</i>	<i>No Preference</i> (<i>n</i> = 134)		<i>Prefer Female</i> (<i>n</i> = 92)		<i>Prefer Male</i> (<i>n</i> = 9)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Feminist Identity Scale</i>						
Passive acceptance	3.18	.75	3.03	.84	3.58	.98
Revelation	2.79	.59	2.95	.56	2.96	.50
Embeddedness-emanation	2.96	.65	3.20	.71	3.02	.96
Synthesis	3.89	.50	3.83	.52	4.10	.47
<i>Feminist Identity Development Scale</i>						
Passive acceptance	2.45	.61	2.39	.62	2.67	.49
Revelation	2.94	.65	3.21	.51	3.17	.60
Embeddedness-emanation	2.66	.61	2.94	.58	2.94	.78
Synthesis	3.97	.54	3.86	.51	3.89	.55
Active commitment	3.14	.56	3.22	.54	3.19	.77
<i>Feminist Identity Composite</i>						
Passive acceptance	2.71	.71	2.63	.74	3.11	.44
Revelation	2.57	.70	2.77	.61	2.60	.45
Embeddedness-emanation	2.95	.82	3.17	.98	2.92	1.18
Synthesis	4.27	.58	4.19	.62	4.53	.58
Active commitment	3.38	.56	3.42	.52	3.48	.68

the time (2 of 18 times; placed in R both times); PA items were not misclassified.

For the FIDS, hit rates ranged from 87% to 100% (*Mdn* = 100%; *M* = 96%), with AC items misplaced 17% of the time (4 of 24 times placed in S) and an S item misplaced 7% of the time (one of 15 times placed in AC). These items were misplaced by the same judge. No misclassifications occurred for PA, R, and EE items.

For the FIC, hit rates ranged from 73% to 94% (*Mdn* = 91%; *M* = 88%). Across the three judges, S items were misplaced 27% of the time (4 of 15 times; placed in EE twice and AC twice), R items were misplaced 17% of the time (4 of 24 times; placed in EE twice and S twice), and AC items were misplaced 15% of the time (4 of 27 times placed in S).

Structural validity. Several scholars have suggested that indicator per factor (p/f) ratios influence fit index values obtained in confirmatory factor analysis (CFA) (Ding, Velicer, & Harlow, 1995; Marsh, Hau, Balla, & Grayson, 1998; Williams & Holahan, 1994). To reduce the possibility that differences in p/f ratios resulted in different model fits across instruments, we followed Hall, Snell, and Foust's (1999) recommendation and rationally derived

“parcels” of items. For each instrument, item content for each subscale was examined, and items that seemed to share a secondary construct were grouped together. Further, to ensure that the CFA results were not idiosyncratic to the parcels used, for each instrument two separate sets of rationally derived parcels were developed. In each case, 3 to 4 parcels per factor/subscale were developed for each instrument. Thus, two separate CFAs were conducted for each instrument (i.e., one CFA with each set of parcels). Differences across the two sets of analyses were small. Results from the set of parcels that produced better fit index values are reported here. (Information about item parcels’ item composition and the second set of analyses is available upon request.)

We used LISREL 8.14 (Jöreskog & Sörbom, 1993) to conduct CFAs, one to assess the structural validity of each instrument. A sample size of 200 (Boomsma, 1982) or 5 cases per estimated parameter (Bentler, 1993) is generally sufficient for conducting a CFA. The greatest number of parameters we estimated was 46 (for the FIDS model). Thus, the minimum required sample size would be 230, and our sample size of 240 was sufficient for the following analyses.

For each instrument, parcel covariance matrices were computed using PRELIS 2.14 and used as input. Paths from the latent variables to corresponding item parcels and intercorrelations among the latent variables were freely estimated. Variance of the latent variables was set to 1.0. Because the χ^2 statistic can be distorted by sample size (see Ullman, 1996), we also examined values for the χ^2/df , goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), non-normed fit index (NNFI, also known as Tucker Lewis index [TLI]), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Rough guidelines for values indicative of a well-fitting model are a χ^2 to df ratio ≤ 2 , GFI, AGFI, CFI, and NNFI values $\geq .90$, RMSEA values $\leq .08$, and SRMR values $\leq .05$ (Ullman, 1996).

Fit index values for the three instruments are reported in Table 4, and model summaries are reported in Figures 1 through 3. For the FIS, a proper solution converged in 12 iterations. Parcel loadings on assigned factors ranged from .12 (EE) to .79 (R). Fourteen of 15 parcels loaded significantly on their assigned factor. One parcel had a substantial ($> |.30|$) cross-loading (loading of EE parcel 2 on R = .47). Interrelations among the four factors ranged from $-.13$ to $.52$ (5 out of 6 of these were significant). No fit index values reached recommended cutoffs.

For the FIDS, a proper solution converged in 12 iterations. Parcel loadings on assigned factors ranged from .18 (S) to .84 (S). All parcels loaded significantly on their assigned factors. Seven of 19 parcels had substantial ($> |.30|$) cross-loadings (loading of R parcel 2 on EE = .31, on AC = .41; loading of R

TABLE 4: Chi-Square Statistic and Fit Index Values for the FIS, FIDS, and FIC

Instrument	χ^2	df	χ^2/df	GFI	AGFI	NNFI	CFI	RMSEA	SRMR
FIS	271.47*	84	3.23	.87	.81	.77	.81	.097	.098
FIDS	311.57*	142	2.19	.89	.85	.84	.87	.071	.085
FIC	263.37*	125	2.11	.89	.85	.87	.89	.068	.067

NOTE: $N = 240$. FIS = Feminist Identity Scale; FIDS = Feminist Identity Development Scale; FIC = Feminist Identity Composite; GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index; NNFI = Non-Normed Fit Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.
* $p < .05$.

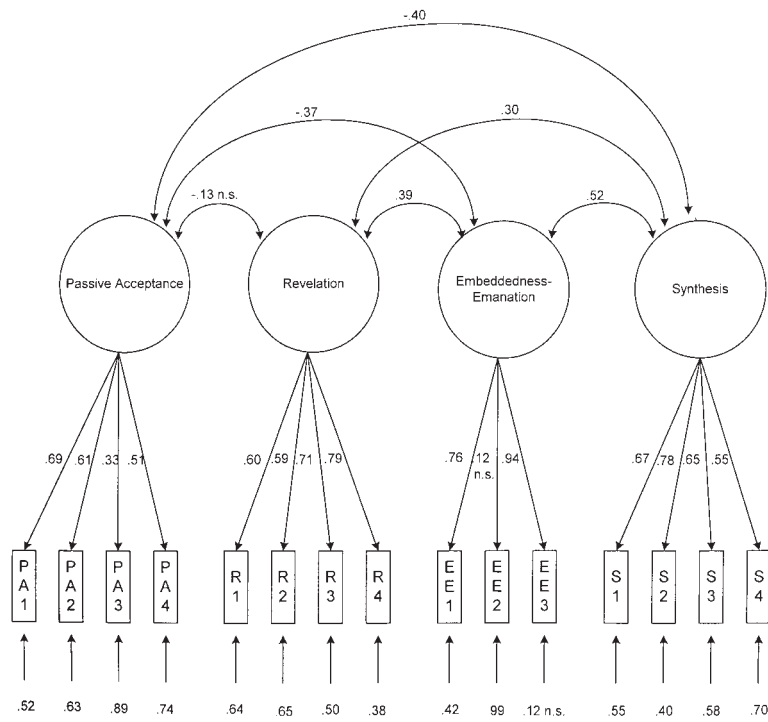


Figure 1. Model for the Feminist Identity Scale (FIS).

parcel 3 on AC = $-.35$; loading of EE parcel 1 on AC = $.32$; loading of EE parcel 3 on AC = $-.44$; loading of AC parcel 3 on PA = $-.35$, on R = $-.39$), but none of these cross-loadings was greater than the loading of the parcels on

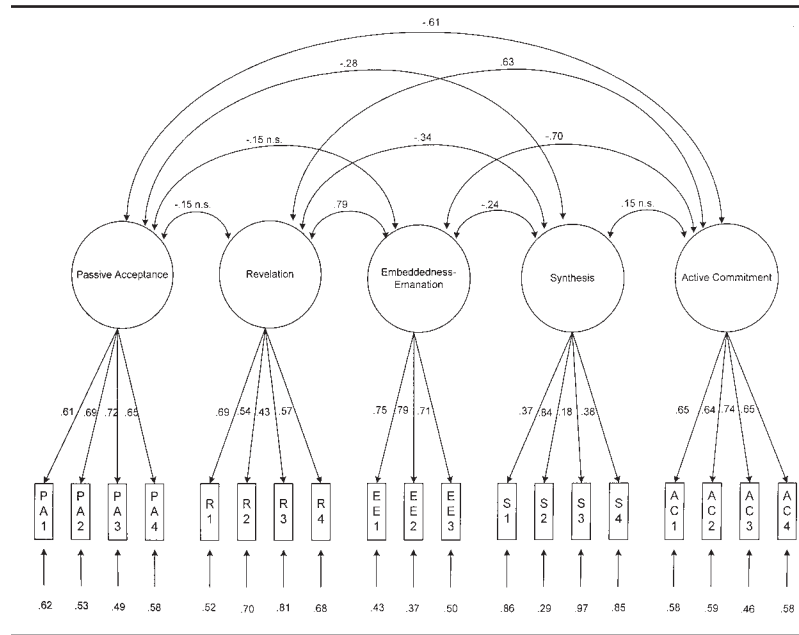


Figure 2. Model for the Feminist Identity Development Scale (FIDS).

their intended factors. Interrelations among the five factors ranged from .15 to .79 (7 of 10 of these relations were significant). All fit index values except the RMSEA approached recommended cutoffs.

For the FIC, a proper solution converged in 11 iterations. Parcel loadings on assigned factors ranged from .27 (S) to .84 (EE). All parcels loaded significantly on their assigned factors. Three of the 18 parcels cross-loaded substantially ($>|.30|$) (loading of S parcel 1 on AC = .40, loading of S parcel 2 on AC = -.40, loading of AC parcel 4 on S = .47), but none of these cross-loadings were greater than the parcel's loading on its assigned factor. Interrelations among the five factors ranged from -.03 to .64 (8 of 10 of these relations were significant). With the exception of the RMSEA, fit index values approached but did not reach recommended cutoffs.

In general, data from all three instruments produced fit index values that were slightly below most conventionally accepted cutoffs, but the FIDS and FIC produced slightly better model-data fits. Further, examination of parcels' factor loadings and cross-loadings suggests an overall "cleaner" fit for the FIDS and FIC (i.e., all parcels loaded significantly on assigned factors with no cross-loadings exceeding the value of the assigned factor loading).

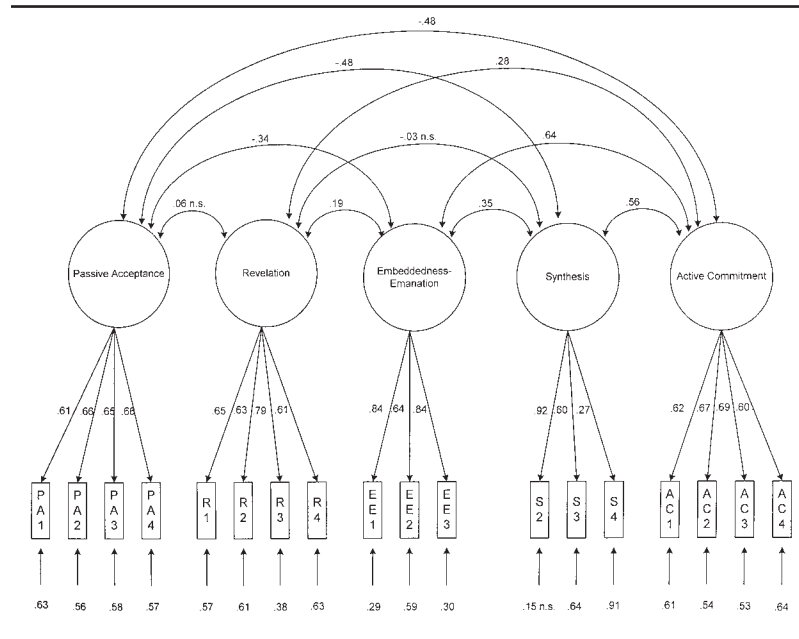


Figure 3. Model for the Feminist Identity Composite (FIC).

DISCUSSION

This study advanced previous research on measurement of Downing and Roush's (1985) model by assessing the psychometric properties of the FIS, FIDS, and FIC using a single sample to reduce sample-specific influences on comparisons. In response to previous researchers' exclusive reliance on exploratory factor analyses (except for Fischer et al.'s [2000] analysis of the FIC), this study employed confirmatory factor analyses, which may be a more appropriate technique when testing theory with regard to factor structure (Ullman, 1996).

Consistent with previous research (e.g., Fischer et al., 2000; Gerstmann & Kramer, 1997; Juntunen et al., 1994), this data suggested problematic internal consistency reliabilities for the FIS and FIDS. The unacceptable alpha reliabilities for FIS PA, FIS EE, FIDS R, and FIDS S subscales may lead to underprediction of the links of their corresponding feminist identity development attitudes to other variables of interest. Thus, these subscales should be used cautiously (if at all) in research and practice. Consistent with Fischer et al.'s (2000) findings, however, FIC subscales may represent more reliably relations between feminist identity development and other variables of inter-

est. Indeed, in the present assessment of convergent validity, subscales with the lowest alphas (e.g., FIS PA, FIDS S) produced the greatest number of nonsignificant correlations.

Stability of FIS and FIDS subscales generally was slightly lower than had been reported in prior research (i.e., Gerstmann & Kramer, 1997; Rickard, 1989), but the stability of the FIDS AC was substantially lower ($r = .38$) than the estimate provided by Gerstmann and Kramer (1997) ($r = .85$). The FIC subscale stabilities were similar to those of the FIS and FIDS. These modest results are puzzling, given the short interval, but may reflect especially dynamic attitudes in this sample. Indeed, Gerstmann and Kramer's inclusion of women's studies students in their sample and Rickard's (1989) use of students in a psychology of women course may have introduced a confound; women who enroll in such classes may have more developed and stable attitudes about feminist identity (especially AC) than did the present sample.

Consistent with prior research (e.g., Fischer et al., 2000), a strength of all three instruments was that none of the subscales related significantly to social desirability. Thus, when responses remained anonymous, none of the instruments elicited socially desirable responding. This may not generalize to conditions where responses are not anonymous (e.g., when used for assessment or intervention in counseling); future researchers are encouraged to examine this.

Correlations between the S and AC subscales of each of the three instruments and self-esteem generally were consistent with expectations (i.e., for each instrument, S or AC scores were related positively to self-esteem), but each instrument also yielded some unexpected nonsignificant relations. Particularly, the notion that PA relates negatively to self-esteem (Downing & Roush, 1985) was not supported when PA was measured with the FIS or FIC. Also, the positive relation of S to self-esteem (McNamara & Rickard, 1989) was not supported when S was measured by the FIDS. These results suggest that these subscales should be used with caution—although the lack of relations between self-esteem and EE or R may reflect accurately the attitudinal flux associated with these stages rather than psychometric problems.

Therapist preference data were somewhat consistent for the FIDS but less consistent for the FIS and FIC. According to Downing and Roush (1985), women who endorse R and EE attitudes tend to idealize and prefer women over men. Indeed, FIDS scores suggested that women who preferred a female therapist scored higher on R and EE. The nonsignificant FIS and FIC findings raise questions, however, with regard to whether FIS R, FIC R, and FIC EE capture R and EE as theorized by Downing and Roush (1985). It is interesting that although few women (only 4%) preferred a male therapist, the descriptive data were consistent with prediction (i.e., these women had higher PA scores than did women who preferred a female therapist or had no prefer-

ence). Contrary to McNamara and Rickard's (1989) predictions, none of the instruments supported the notion that women with no preference for therapists' gender score higher on S and AC. The majority of our participants (57%), however, regardless of feminist identity attitudes, reported no preference, suggesting that other characteristics (e.g., therapists' feminist identity attitudes) may be more important to potential clients; this possibility warrants future research.

Correlations between the feminist identity development subscales of each of the three instruments and reports of perceived sexist events generally were supportive of the convergent validity of all instruments. For example, consistent with the denial of sexism characteristic of PA, PA subscales of each of the instruments were related negatively or nonsignificantly to reports of sexist events. Also, consistent with vigilance concerning sexism associated with R, all R subscales correlated highest with reports of sexist events. Thus, the present data provide evidence of convergent validity for the PA and R subscales of the instruments. Contrary to expectation, however, the FIDS S subscale, perhaps due to its low internal consistency, did not correlate significantly with reports of sexist events. In contrast, FIS S correlated significantly and positively with each dimension of reports of sexist events. These data, and the fact that FIS S correlated slightly higher with FIDS AC than with FIDS S, suggest that the S subscales of the FIS and the FIDS may not be measuring the same construct. Consistent with Rickard's (1989) thinking, FIS S may measure AC, or FIDS S may not be a cohesive and stable measure of S.

The fact that within each instrument, the intercorrelations among subscales that measure nonadjacent feminist identity development stages were sometimes higher than those for adjacent stages raises questions with regard to the status of Downing and Roush's (1985) model as a linear and sequential process and/or the status of the FIS, FIDS, and FIC as measures of such a process. The lack of direct tests of the developmental aspect of Downing and Roush's model (e.g., longitudinal studies of feminist identity development), and concerns raised by the current data, caution against using scores on feminist identity development instruments to assign women to stages. We encourage researchers to use continuous scores on the FIS, FIDS, and FIC subscales to assess women's attitudes about the feminist identity development stages. Such use of these instruments is consistent with previous suggestions (Fischer & Good, 1994) and researchers' use of measures of racial identity development (e.g., Parham & Helms, 1981, 1985).

Although the hit rates obtained in this study generally were consistent with those reported by Fischer et al. (2000) for the FIS and FIDS, questions about subscale content emerge from the judges' ratings; improvement of the content of the FIC and FIS subscales may be a fruitful area for research. The pattern of the data suggested that FIS items were misclassified more fre-

quently than were FIDS and FIC items, and some FIS and FIC misclassifications were unexpected (e.g., FIS and FIC R items placed in S, FIS S item placed in PA). Relatedly, the fact that the FIS does not assess AC, based on Rickard's (1989) assumption that AC is a manifestation of the S stage, may be problematic without data to suggest that her assumption is warranted. Indeed, the judges' ratings as well as the parcel loadings in the confirmatory factor analyses for the FIDS and FIC suggest that S and AC are related but distinct constructs. Research is needed to examine the contention that AC is a behavioral manifestation of S.

Despite questions concerning content validity, the trend in the confirmatory factor analysis results provides tentative support for the FIC, and perhaps FIDS, with less support for the FIS. The pattern of parcel loadings and cross-loadings for the FIS raised questions about the FIS EE. Only parcels of items reflecting interest in women artists, musicians, writers, and women's studies loaded significantly on FIS EE (these items also make up FIC EE). These items do not capture fully Downing and Roush's (1985) conceptualization of EE. Consistent with Fischer et al.'s (2000) suggestion, we encourage development of new items for the FIS and FIC EE.

Somewhat contrary to these findings, Fischer et al. (2000) reported an excellent data-model fit for the FIC (χ^2 to *df* ratio = 1.94; GFI = .96; NNFI = .95; CFI = .96; SRMR = .046). This inconsistency in findings may exist because Fischer et al. used fewer indicators (i.e., 2-3) than were used in this study (i.e., 3-4). Fewer indicators per factor may result in better fit index values (e.g., Ding et al., 1995), but several scholars have suggested that a minimum of three indicators per factor should be used in confirmatory factor analyses (e.g., Ding et al., 1995; Hall et al., 1999). Future researchers should consider this and examine further the effect of the number of indicators per factor on the model-data fit of measures of feminist identity attitudes.

Limitations and Implications for Future Research

We encourage researchers and practitioners to continue building on the strengths of the FIS, FIDS, and FIC to improve assessment of feminist identity development. The most specific and pragmatic area for future development may be improvement of the FIC's EE subscale. The FIC outperformed the FIS and FIDS in terms of internal consistency reliability and structural validity, but it inadequately assessed EE. As the goal of this study was to compare existing feminist identity development instruments to identify areas for improvement, we did not develop or examine additional items for subscales; this seems a potentially fruitful area for future work.

Another important area for future research is examination of the generalizability of these findings to women from diverse backgrounds. This

study did not address specifically the applicability or utility of the construct of feminist identity development and its operationalizations to racial/ethnic or other minority women. Indeed, the relatively homogeneous racial/ethnic and class composition of this sample, and the fact that most women were young and either had a college degree or were enrolled in classes working toward one, limit the generalizability of these results. Theory and assessment of feminist identity development will be informed by studies that examine the reliability and structural validity of measures of feminist identity development for women of diverse ages and backgrounds. Further, the modest response rate for faculty/staff women is a concern, especially given that we do not know if participants were similar to nonparticipants. Perhaps qualitative studies to explore diverse women's experiences of feminist identity are needed to inform theory and research.

Our examination of the links between feminist identity development attitudes and therapy-related variables was limited to nonclients' self-reported preferences. Examining links of clients' and therapists' feminist identity attitudes to therapeutic process and outcome (e.g., working alliance) may further illuminate the role of feminist identity development in therapy.

Finally, the order of presentation of instruments may have affected participants' responses. Perhaps fatigue or boredom led to careless responding to instruments presented later in the survey packet. The consistency of our results with previous findings fails to support this as a problem, yet our results must be interpreted in light of this methodological limitation.

Summary and Recommendations

Overall, the present data suggested that each of the three current measures of feminist identity development has strengths and limitations. Given these mixed results, we encourage researchers and practitioners to consider their needs in assessing feminist identity development and to select the instrument (or combination of instruments) that best meets those needs. For example, if interested in assessing parts of the feminist identity development model, one may choose to use the instrument with the strongest psychometric properties for the subscales of interest. If interested in assessing the entire model, we recommend the FIC, because it performed comparatively better in terms of internal consistency reliability and structural validity than the FIS and FIDS. Alternatively, one may choose to administer the FIS and FIDS and use the best subscales to assess the full model (i.e., FIDS PA, FIS R, FIDS EE, FIS S, FIDS AC).

A more serious issue is whether corresponding subscales across instruments measure different aspects of their intended feminist identity development stage (e.g., FIDS PA is related to lower self-esteem but FIS and FIC PA

are not). Relatedly, inadequate assessment of EE by the FIS and FIC raises concerns. The present data do not address whether the observed psychometric limitations originate in problematic measurement or inaccurate theory. Clarifying this issue is essential in advancing understanding of feminist identity development.

REFERENCES

- Bargad, A., & Hyde, J. S. (1991). Women's studies: A study of feminist identity development in women. *Psychology of Women Quarterly, 15*, 181-201.
- Bentler, P. M. (1993). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction (Part 1)* (pp. 149-173). Amsterdam: Elsevier Science.
- Cash, T. F., Ancis, J. R., & Strachan, M. D. (1997). Gender attitudes, feminist identity, and body images among college women. *Sex Roles, 36*, 433-447.
- Ding, L., Velicer, W. F., & Harlow, L. 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.
- Downing, N. E., & Roush, K. L. (1985). From passive acceptance to active commitment: A model of feminist identity development for women. *The Counseling Psychologist, 13*, 695-709.
- Fischer, A. R., & Good, G. E. (1994). Gender, self, and others: Perceptions of the campus environment. *Journal of Counseling Psychology, 41*, 343-355.
- Fischer, A. R., Tokar, D. M., Mergl, M. M., Good, G. E., Hill, M. S., & Blum, S. A. (2000). Assessing women's feminist identity development: Studies of convergent, discriminant, and structural validity. *Psychology of Women Quarterly, 24*, 15-29.
- Gerstmann, E. A., & Kramer, D. A. (1997). Feminist identity development: Psychometric analyses of two feminist identity scales. *Sex Roles, 36*, 327-348.
- 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.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago: Scientific Software.
- Juntunen, C. L., Atkinson, D. R., Reyes, C., & Gutierrez, M. (1994). Feminist identity and feminist therapy behaviors of women psychotherapists. *Psychotherapy, 31*, 327-333.
- Klonoff, E. A., & Landrine, H. (1995). The Schedule of Sexist Events: A measure of lifetime and recent sexist discrimination in women's lives. *Psychology of Women Quarterly, 19*, 439-472.
- Landrine, H., & Klonoff, E. A. (1997). *Discrimination against women: Prevalence, consequences, remedies*. Thousand Oaks, CA: Sage.
- 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.
- McNamara, K., & Rickard, K. M. (1989). Feminist identity development: Implications for feminist therapy with women. *Journal of Counseling & Development, 68*, 184-189.

- Moradi, B., & Subich, L. M. (2002). Feminist identity development measures: Comparing the psychometrics of three instruments. *The Counseling Psychologist, 30*, 66-86.
- Moradi, B., Subich, L. M., & Phillips, J. C. (2002). Revisiting feminist identity development theory, research, and practice. *The Counseling Psychologist, 30*, 6-43.
- Ng, S. H., Dunne, M., & Cataldo, M. (1995). Feminist identities and preferred strategies for advancing women's positive self-concept. *Journal of Social Psychology, 135*, 561-572.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Parham, T. A., & Helms, J. E. (1981). The influence of Black students' racial identity attitudes on preference for counselor's race. *Journal of Counseling Psychology, 28*, 250-256.
- Parham, T. A., & Helms, J. E. (1985). Relation of racial identity attitudes to self-actualization and affective states of Black students. *Journal of Counseling Psychology, 32*, 431-440.
- Paulhus, D. L. (1994). *Balanced inventory of desirable responding: Reference manual for BIDR version 6*. Vancouver: University of British Columbia.
- Rickard, K. M. (1989). The relationship of self-monitored dating behavior to level of feminist identity on the Feminist Identity Scale. *Sex Roles, 20*, 213-226.
- Rickard, K. M. (1990). The effect of feminist identity level on gender prejudice toward artists' illustrations. *Journal of Research in Personality, 24*, 145-162.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (3rd ed., pp. 709-811). New York: HarperCollins.
- Williams, L. J., & Holahan, P. J. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work? *Structural Equation Modeling, 1*, 161-189.
- Wylie, R. C. (1989). *Measures of self-concept*. Lincoln: University of Nebraska Press.