Journal of Applied Psychology
1996, Vol. 81, No. 4, 400-410

Copyright 1996 by the American Psychological Association, Inc.
0021-9010/96/$3.00

Development and Validation of Work–Family Conflict and Family–Work Conflict Scales

Richard G. Netemeyer
Louisiana State University

James S. Boles
Georgia State University

Robert McMurrian
Louisiana State University

Researchers report on a 3-sample study that developed and validated short, self-report scales of work–family conflict (WFC) and family–work conflict (FWC). Using conceptualizations consistent with the current literature, the researchers offer content domains and definitions of the constructs. Advocated procedures were used to develop the scales and test dimensionality and internal consistency. Estimates of construct validity are presented by relating the scales to 16 other on- and off-job constructs. Mean-level difference tests between WFC and FWC also provide evidence of validity.

A widely studied topic in organizational behavior is the conflict between work and family. In recent years, several studies have advanced our understanding of how work affects family life and vice versa (Frone, Russell, & Cooper, 1992; Greenhaus & Beutell, 1985; O'Driscoll, Ilgen, & Hildreth, 1992). Of importance to the study of work–family relations is construct measurement. Although a conceptual distinction between work conflicting with the family and family conflicting with work has been made, most research has assessed only work–family conflict (WFC). Furthermore, operationalization of WFC has varied widely from study to study, limiting the ability to generalize about the effects of WFC. The purpose of this article is to develop and validate short, self-report measures of WFC and family–work conflict (FWC).

Background and Review

Two important focal points of adult life are family and work. However, the role expectations of these two domains are not always compatible, creating conflicts between work and family life. These conflicts are related to outcomes such as job dissatisfaction, job burnout, and turnover (Burke, 1988; Frone et al., 1992; Greenhaus, 1988; Pleck, Staines, & Lang, 1980), as well as to outcomes related to psychological distress (e.g., depression), and life and marital dissatisfaction (Greenhaus & Beutell, 1985; Gutek, Searle, & Klepa, 1991; Voydanoff, 1988). Given the increase in dual-earner families, single-parent families, and families with elder-care duties, these outcomes are likely to be even more pronounced in the future. Furthermore, there is mounting evidence that WFC and FWC are related to work productivity and financial costs incurred by an organization (Cascio, 1991).

Although many studies have examined the relationships between WFC and other variables, their measures have varied widely. For example, WFC has been operationalized with single-item measures that may lack reliability (Rice, Frone, & McFarlin, 1992; Voydanoff, 1988) and with lengthy measures that are possibly cumbersome to respondents (Burke, 1988; Burke, Weir, & Duwors, 1979). Several studies have combined WFC and FWC scales into a single measure, ignoring the conceptual distinction between the two constructs (Cooke & Rousseau, 1984; Kopelman, Greenhaus, & Connolly, 1983; Thomas & Ganster, 1995; Wiley, 1987). Other studies have offered separate WFC and FWC measures, but have used items that reflect potential outcomes of the constructs rather than their content domain, that is, items that assess somatic, physical, and mental symptoms that are due to WFC, FWC, or both (Bedeian, Burke, & Moffett, 1988; O'Driscoll et al. 1992; Parasuraman, Greenhaus, Rabinowitz, Bedeian, & Mossholder,

Richard G. Netemeyer and Robert McMurrian, Department of Marketing, Louisiana State University; James S. Boles, Department of Marketing, Georgia State University.

We thank Art Bedeian for judging items and offering constructive comments on an earlier version of this article. We also thank Cheryl Adkins and Roger Griffeth for judging items and the Georgia State University Marketing Roundtable for supporting this research.

Correspondence concerning this article should be addressed to Richard G. Netemeyer, 3126B CEBA Building, Department of Marketing, Louisiana State University, Baton Rouge, Louisiana 70803. Electronic mail may be sent via Internet to mknete@lsuvm.sncc.lsu.edu.

1989). Still other studies have used WFC measures that simply have not been subjected to rigorous scale-development procedures (Frone et al., 1992; Gutek et al., 1991; Judge, Boudreau, & Bretz, 1994). Similar observations hold for FWC. To better understand the interplay between these two constructs and their antecedents and outcomes, researchers require sound measures.

WFC and FWC Defined

The conceptual approach taken in the present research is based on the premise that WFC and FWC are distinct but related forms of interrole conflict (Greenhaus & Beutell, 1985; Kahn, 1981; Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964; Pleck et al., 1980). Interrole conflict has been viewed as a form of conflict in which "role pressures associated with membership in one organization are in conflict with pressures stemming from membership in other groups" (Kahn et al., 1964, p. 20). From work–family and family–work perspectives, this type of conflict reflects the degree to which role responsibilities from the work and family domains are incompatible, that is, "participation in the work (family) role is made more difficult by virtue of participation in the family (work) role" (Greenhaus & Beutell, 1985, p. 77). As such, the demands of one role make performance of the other role more difficult (Katz & Kahn, 1978).

Although several sources of WFC and FWC have been identified, most researchers agree that the general demands of a role, the time devoted to a given role, and the strain produced by a given role are domain elements of WFC and FWC (Bachrach, Bamberger, & Conley, 1991; Cooke & Rousseau, 1984; Greenhaus, 1988; Greenhaus & Beutell, 1985; Gutek et al., 1991; Kahn & Byosiere, 1992; Kahn et al., 1964; Pleck et al., 1980; Voydanoff, 1988). The general demands of a role refer to the *responsibilities, requirements, expectations, duties,* and *commitments* associated with a given role. (These terms have been used interchangeably throughout the literature.)

Time-based conflict occurs when the amount of time devoted to the work (family) role interferes with performing family- (work-)related responsibilities. Specifically, excessive work (family) time conflicts may make it difficult to comply with family (work) responsibilities. Strain-based conflict occurs when strain created by the work (family) role interferes with performing family (work) responsibilities. For example, irritability and anxiety created by work interfere with performing family duties and vice versa.

As such, we used the following definitions to guide our scale development. WFC is a form of interrole conflict in which the general demands of, time devoted to, and strain created by the job interfere with performing family-related responsibilities. FWC is a form of interrole conflict in which the general demands of, time devoted to, and strain created by the family interfere with performing work-related responsibilities.

Relations With Other Variables and Mean-Level Difference

Investigating the construct validity of the WFC and FWC scales developed in this study called for a number of predictions to be advanced. These predictions pertain to the relationships between WFC and FWC and other on- and off-job constructs, as well as mean-level differences between WFC and FWC scores. The following sections summarize our predictions and the rationale for our predictions.

On-Job Constructs

Research suggests there is an inverse relationship between organizational commitment and WFC and FWC (O'Driscoll et al., 1992; Greenhaus & Beutell, 1985) and between job satisfaction and WFC and FWC (Frone et al., 1992; Rice et al., 1992). Thus, negative correlations between the WFC and FWC scales and organizational commitment and job satisfaction are predicted.

It has been suggested that WFC and FWC should be positively associated with job burnout, job tension, job role conflict, and job role ambiguity (Bedeian et al., 1988; Frone et al., 1992; Maslach & Jackson, 1981). Thus, positive correlations between these variables and WFC and FWC are predicted. Prior research also indicates that WFC is more strongly related to job burnout and job tension than FWC (O'Driscoll et al., 1992; Frone et al., 1992; Judge et al., 1994; Maslach & Jackson, 1981). Given these findings, we predict that WFC correlates more strongly with job burnout and job tension than with FWC.

Research shows that intention-to-leave-an-organization and search-for-another-job are positively related to WFC and FWC (Burke, 1988). Therefore, we predict positive correlations between intention-to-leave-an-organization and search-for-another-job and the WFC and FWC measures. Finally, the numbers of hours worked per week should be more highly correlated with WFC than with FWC (Gutek et al., 1991).

Because salespeople were the respondents for the third sample in this study, the following predictions were advanced. Sales self-efficacy influences expectations about one's ability to perform a job successfully (Bandura, 1986). Thus, it is predicted that FWC should have a negative correlation with self-efficacy. The sales literature also suggests an inverse relation between FWC and sales performance (Behrman & Perreault, 1984). Thus, it is predicted that FWC should have a negative correlation with sales performance. To our knowledge, no empirical research has assessed the differences in relationships be-

tween WFC and FWC and self-efficacy and sales performance. However, it seems reasonable to suggest that FWC should be more strongly related to one's perceived ability to perform a job and one's job performance than WFC. Thus, the correlations between FWC and self-efficacy, and FWC and sales performance should be stronger than the correlations between WFC and self-efficacy and WFC and sales performance.

Off-Job Constructs

For the variables of life satisfaction, relationship satisfaction, and relationship agreement, inverse relationships with FWC and WFC have been suggested (Judge et al., 1994; Parasuraman et al., 1989). As such, negative correlations between WFC and FWC and these variables are predicted. Because both WFC and FWC are related to negative physical symptoms and depression (Burke, 1988; Frone et al., 1992; Kemery, Mossholder, & Bedeian, 1987), physical symptomology and depression are predicted to be positively correlated with WFC and FWC. Finally, persons with more children (at home) must adjust their demands, time, and emotions between the work and home setting more than persons who have few or no children. Thus, the number of children at home should be positively correlated with WFC and FWC.

Mean-Level Difference Between WFC and FWC

Most workers report that family is more important than work, and research indicates that WFC is greater than FWC (Gutek, Repetti, & Silver, 1988; Gutek et al., 1991; Judge et al., 1994). Therefore, we predict the WFC scale should exhibit a higher mean score than the FWC scale.

Overview of Scale Development

The procedures we used to develop the WFC and FWC scales closely adhere to those described in the psychometric literature (Cortina, 1993; DeVillis, 1991; Robinson, Shaver, & Wrightsman, 1991; Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993). After construct definition, these procedures included item generation and judging, measure purification, examination of dimensionality and internal consistency, measurement invariance testing, and construct validity assessment.

Method

Item Generation and Judging

A large pool of statements (*items*) was generated to reflect the WFC and FWC conceptualizations. Items were culled from previously published sources that we felt reflected the domains of the constructs (Bachrach et al., 1991; Bedeian et al., 1988; Burke et al., 1979; O'Driscoll et al., 1992; Frone et al., 1992; Kopelman et al., 1983; Wiley, 1987). Several of these items required slight wording modifications to fit the Likert format used in our approach, and most considered a *partner* to reflect nonmarried or nontraditional relationships (i.e., a significant other). We generated other items such that a total of 110 items served as the initial pool of statements. Of these, 18 items each were generated to reflect general demand WFC and general demand FWC. Twenty and 19 items reflected time- and strain-based WFC. Nineteen and 16 items reflected time- and strain-based FWC. About one third of all the statements were reverse worded.

To reduce the pool of items to a more manageable number, a panel of four faculty members from other universities judged the items for representativeness. These judges whose primary area of research was organizational behavior were from departments of management at two major state universities. The judges were given the construct definitions and were asked to evaluate each item as *very representative, somewhat representative,* or *not representative* of the definitions. Interrater reliability coefficients were constructed. The formula used was a variation of Cohen's kappa where the coefficient ranges from a low of 0 to a high of 1 (Jones, Johnson, Butler, & Main, 1983). For all four judges simultaneously, the value of this coefficient was only .52. However, when two judges at a time were considered, the values were higher, ranging from .63 to .79. As a result, only those items that all four judges classified the same and were rated at least *somewhat representative* of the construct definition were retained.

Further exploratory analyses reduced the pool to 43 items for the samples that follow (i.e., 7, 8, and 7 items for general demand and time- and strain-based WFC, respectively, 22 from a total of 22; and 8, 7, and 6 items for general demand and time- and strain-based FWC, respectively, from a total of 21). Following are descriptions of the samples and measures used in scale development and validation.

Sample 1

Respondents and procedure. Questionnaires were sent to elementary and high-school teachers and administrators in a large southeastern city. A cover letter assured them of the confidentiality and anonymity of their responses. Of the 224 mailed, 182 were returned for an effective response rate of 81%. Of the 182 respondents, 128 were women, the median age was 43, 157 were married, and 93 had children living at home.

Measures. Sample 1 responded to the 43 WFC and FWC items. These items were responded to along 7-point *strongly disagree–strongly agree* response scales. Several on- and off-job measures dealing with variables other than WFC and FWC were also included in the questionnaire. The on-job measures were organizational commitment, job satisfaction, job burnout, job tension, job role conflict, job role ambiguity, intention-to-leave-an-organization, search-for-another-job, and number of hours worked per week.

Organizational commitment was measured by a 9-item version of the scale developed by Mowday, Steers, and Porter (1979). Job satisfaction was measured with five items culled from various sources (Price & Mueller, 1986; Staines & Pleck, 1984). Example items included, "I feel fairly well satisfied with my present job" and "All things considered (i.e., pay, promo-

tion, supervisors, or co-workers), how satisfied are you with your present job?" All satisfaction items were responded to along 7-point response scales.

Job burnout was measured with the Maslach Burnout Inventory (MBI; (Maslach & Jackson, 1981). Job tension was measured with seven items from the anxiety–stress scale developed by House and Rizzo (1972). Job role conflict and role ambiguity were measured with the six- and eight-item scales developed by Rizzo, House, and Lirtzman (1970). Intention-to-leave-an-organization and search-for-another-job were each measured with five-item scales used in previous research (Bluedorn, 1982; Hendrix, Nestor, & Troxler, 1985). Examples of intention-to-leave items were, "I intend to leave my position during the next 12 months" and "I intend to quit my present job." Items used to measure search-for-another-job included, "I have searched for an alternative job since I joined this organization" and "I am actively seeking a job or role (an activity other than my present job)." Intention-to-leave and search-for-another-job items were responded to along 5-point *strongly agree–strongly disagree* response scales. One single-item question, "How many hours per week, on average, do you work on your job (whether at the workplace or at home)?" was included in the questionnaire.

The off-job constructs of life satisfaction, relationship satisfaction, and the level of relationship agreement on key issues were measured as follows. Life satisfaction was measured by a 15-item scale that assesses general happiness with life (Quinn & Staines, 1979). Relationship satisfaction and relationship agreement were measured by 3 and 8 items, respectively, from the Locke and Wallace (1959) Marital Adjustment Test. The items were modified to include persons currently married and those involved in an emotionally binding relationship (i.e., persons with a significant other).

Sample 2

Respondents and procedure. Questionnaires were sent to small business owners in a large southeastern city. A cover letter assured them of the confidentiality and anonymity of their responses. Of the 298 questionnaires mailed, 162 were returned for a response rate of 54%. The median age of respondents was 45 years, 96 were men, 130 were married, and 65 had children living at home.

Measures. With the exception of the role conflict, role ambiguity, and organizational commitment scales, Sample 2 responded to the same measures as Sample 1. Two new measures, physical symptomology and depression, were added to the questionnaire sent to Sample 2. Physical symptomology was assessed with the 55-item checklist derived from a scale developed by Pennebaker (1982). Respondents were asked to indicate the frequency with which each negative physical symptom had been experienced (*never experienced* to *experience it more than once a week*) in the past 6 months. The Automatic Thoughts Questionnaire (ATQ) was also included. The ATQ is a 30-item scale used as a measure of depressive thoughts and behaviors (Hollon & Kendall, 1980).

Sample 3

Respondents and procedure. Questionnaires and postage-paid return envelopes were mailed to real estate salespeople in a large southeastern city. A cover letter assured the salespeople of the confidentiality and anonymity of their responses. Of the 700 questionnaires mailed, 186 were completed for a response rate of 27%. The median age of respondents was 48 years, 142 were women, 148 were married, and 60 had children living at home.

Measures. Sample 3 responded to the same measures as Sample 2. In addition, two new measures, sales self-efficacy and sales performance, were added. The self-efficacy measure contained eight items adapted from Bandura's (1986) view of self-efficacy and a salesperson measure of self-efficacy (Sujan, Weitz, & Kumar, 1994). Example items included, "I feel I am very capable at the task of selling" and "Overall, I am confident in my ability to perform this job well." Self-rated sales performance was composed of five items rated on 7-point scales of *among the worst in the company* to *among the best in the company* (Brown & Peterson, 1994). Example items included, "How do you rate yourself in terms of the quantity of work (e.g., sales) you achieve?" and "How do you rate yourself in terms of the quality of your performance in regard to customer relations?" For all samples, scores on measures dealing with variables other than WFC and FWC were summed to form indices for each construct.

Results

Measure Purification

Using an iterative confirmatory procedure with LISREL VII (Jöreskog & Sörbom, 1989), we derived the final forms of the WFC and FWC scales. For the first iteration, there was a common pool of 43 items relating to WFC and FWC. These items were specified to a correlated two-factor confirmatory model for each of the three samples—the two factors reflecting a 22-item WFC factor and a 21-item FWC factor.

On the basis of a number of heuristics suggested in the scale development literature (Bagozzi & Yi, 1988; DeVillis, 1991), items were deleted that (a) through inspection of modification indices and expected change values did not load higher on their intended factor (e.g., WFC) than the other factor (FWC); (b) consistently resulted in within-factor correlated measurement error, across-factor correlated measurement error, or both (e.g., exhibited a large number of standardized residuals >2.58 with other items); (c) had completely standardized factor loadings <.50; (d) were highly redundant in terms of wording with other items; and (e) had extremely high completely standardized factor loadings (i.e., >.90). In general, items with extremely high factor loadings were redundant in terms of item wording and resulted in within-factor correlated measurement error as well (Bagozzi & Yi, 1988). These heuristics were applied while maintaining the general demand and time- and strain-based conceptualization. That is, general demand and time- and strain-based items were carried over from iteration to iteration given that they met the heuristic levels set for item retention.

Table 1
Estimates of Fit Indices

Model	χ^2	*df*	GFI	AGFI	CFI	TLI
			Sample 1 (n = 182)			
2 Factor	76.24**	34	.92	.87	.96	.94
1 Factor	369.37**	35	.60	.37	.65	.55
			Sample 2 (n = 162)			
2 Factor	85.47**	34	.90	.84	.93	.91
1 Factor	288.66**	35	.68	.49	.66	.55
			Sample 3 (n = 186)			
2 Factor	70.03**	34	.93	.88	.97	.96
1 Factor	410.72**	35	.61	.38	.65	.54

Note. GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index; TLI = Tucker–Lewis Index.
** $p < .01$.

After the first iteration, 13 WFC and 11 FWC items were retained for the next iteration, and a slightly different set of heuristics was applied. Items were deleted that (a) still exhibited correlated measurement errors, (b) had across-factor loadings relatively equal to within-factor loadings, (c) had completely standardized factor loadings <.60, and (d) reflected redundancy in terms of wording with other items. After the second iteration, 7 WFC and 6 FWC items were retained for the nest iteration. The third iteration resulted in the deletion of 3 more items (on the basis of author judgment in terms of redundancy of item wording). The final five-item forms of the scales are displayed in the Appendix.

Dimensionality and Internal Consistency

Confirmatory factor analysis was used to assess scale dimensionality, discriminant validity, and internal consistency of the final form of the scales (Anderson & Gerbing, 1988; Fornell & Larcker, 1981). Two models were estimated: (a) a two-factor model (i.e., two correlated first-order factors) representing the hypothesized WFC-FWC structure in which the individual items were permitted to load only on their hypothesized factors, with no cross-loadings or correlated measurement errors, and (b) a one-factor model in which all WFC and FWC items were specified to a single factor. The latter model was used for comparison purposes.

The top half of Table 1 presents fit statistics for the two models. Taken as a whole, the fit statistics suggest adequate fit for the two-factor model. Across samples, the goodness-of-fit index (GFI) ranged from .90 to .93, and the adjusted-goodness-of-fit index (AGFI) ranged from .84 to .88. Because it has been suggested that GFI and AGFI may suffer from inconsistencies from sampling characteristics (Bollen, 1989; Hoyle & Panter, 1994), we also report two fit indices that have been viewed as robust to sampling characteristics: the Tucker–Lewis index (TLI) and Bentler's (1990) comparative fit index (CFI). Values in the .90 range have been noted as designating adequate fit for these indices. Table 1 shows that the two-factor model exhibited values above .90 across the samples.

Evidence of internal consistency is provided by construct reliability, coefficient alpha, and average variance extracted estimates (see Table 2). Construct reliability is a LISREL-generated estimate of internal consistency analogous to coefficient alpha (Fornell & Larcker, 1981, Equation 10). As Table 2 shows, the two alpha estimates ranged from .82 to .90. Average variance extracted estimates assessed the amount of variance captured by a construct's measure relative to random measurement error. Average variance extracted estimates of .50 or above provide further evidence of internal consistency for a construct's measure (Fornell & Larcker, 1981, Equation 11). All but one average variance extracted estimate achieved this criterion. In addition, the completely standardized within-factor item loadings ranged from .60 to .89 across the samples. An examination of the modification indices for the two-factor models revealed only two significant estimated changes for cross-loading items (i.e., a value of .50 for a FWC item in Sample 1 and a value of .30 for a WFC item in Sample 3). All other across-factor item loadings were below .30 across the samples.

Tests of discriminant validity were also performed. First, the ϕ estimates (i.e., the completely standardized correlations between WFC and FWC) were .48, .33, and .42 for Samples 1, 2, and 3, respectively. It has been suggested that if the square of the parameter estimate between two constructs (ϕ^2) is less than the average variance extracted between the two constructs, discriminant validity is supported (Fornell & Larcker, 1981). This criterion was met by all of the samples in the study.

Also, the one-factor model was compared with the hypothesized two-factor model. If the chi-square fit of the two-factor model is better than the fit of the one-factor model, evidence of discriminant validity among factors exists (Anderson & Gerbing, 1988). For Sample 1, the fit of the two-factor model was better than the fit of the one-factor model, $\chi^2(1, N = 182) = 293.13$, $p < .01$. For Sample 2, the fit of the two-factor model was better than the fit of the one-factor model, $\chi^2(1, N = 162) = 203.19$, $p < .01$. For Sample 3, the fit of the two-factor model was better than the fit of the one-factor model, $\chi^2(1, N = 186) = 340.69$, $p < .01$.

Measurement Invariance Tests

Multiple-group measurement invariance tests with LISREL VII were performed on the WFC and FWC scales. When parallel data exist across groups, multiple-group

Table 2
Internal Consistency Estimates for Work–Family Conflict (WFC) and Family–Work Conflict (FWC)

	WFC			FWC		
Sample	Construct α	Coefficient α	Average	Construct α	Coefficient α	Average
1	.88	.88	.60	.87	.86	.58
2	.89	.89	.60	.82	.83	.48
3	.88	.88	.59	.90	.89	.64

Note. Construct α = construct reliability; average = average variance extracted estimate.

analysis offers a powerful test of the equivalence of factor solutions across samples because it rigorously assesses measurement properties (Bagozzi & Yi, 1988; Bollen, 1989; Marsh, 1995; Marsh & Hocevar, 1985).

In general, models of invariance are tested hierarchically, where the hierarchy begins with the most restrictive model—a pattern of fixed and nonfixed parameters invariant across groups. If this model shows reasonable fit, it is used as a baseline for comparison with subsequent models in the hierarchy. Although there is no consensus on the ordering of the subsequent models of invariance, researchers recommend the following hierarchy for measurement models: (a) a model specifying invariant factor loadings across groups; (b) a model specifying invariant factor loadings and invariant factor correlations across groups; and (c) a model specifying invariant factor loadings, invariant factor correlations, and invariant factor variances across groups (Marsh, 1995; Marsh & Hocevar, 1985).

Each sample's parameter specifications were compared simultaneously. Table 3 presents the fit estimates for the models in the invariance hierarchy. The baseline model shows adequate fit as the indices (GFI, CFI, and TLI) are in the .90 range and above. Thus, the model constraining the factor loadings to be invariant across groups was estimated. The difference in fit between this model and the baseline model was $\chi^2(20, N = 530) = 58.52$, $p < .01$, indicating that there is some nonchance lack of invariance. However, it is important to note that statistical tests of invariance have the same limitations as statistical tests for any other confirmatory model. That is, "invariance constraints are a priori false when applied to real data with a sufficiently large sample size" (Marsh, 1995, p. 12). Thus, fit indices should also be used to assess invariance. If the fit indices are adequate, reasonable evidence of parameter invariance exists (Marsh, 1995; Marsh & Hocevar, 1985). As Table 3 indicates, adequate fit was found for the factor loadings invariant model across indices.

The next model estimated was the model that constrained the factor loadings and factor correlations invariant across groups. The difference in fit between this model and the baseline model was $\chi^2(22, N = 530) = 64.19$, $p < .01$. However, levels of fit for the factor loadings and factor correlations invariant model were adequate. The last model estimated was the model that constrained the factor loadings, factor correlations, and factor variances invariant across groups. This model was compared with the baseline model. The difference in fit between the two models was $\chi^2(42, N = 530) = 158.03$, $p < .01$. Although this chi-square difference was relatively large, invariant factor variances are considered the least important in testing measurement property invariance across groups (Bollen, 1989; Marsh, 1995). Further-

Table 3
Tests of Measurement Invariance

Model	χ^2 (N = 530)	df	χ^2_{diff}[a]	df_{diff}[b]	GFI	CFI	TLI
No constraints (baseline model)	231.74**	102			.93	.95	.94
Factor loadings invariant	290.26**	122	58.52**	20	.90	.94	.93
Factor loadings and factor correlations invariant	295.93**	124	64.19**	22	.92	.94	.93
Factor loadings, factor correlations, and factor variances invariant	389.77**	144	158.03**	42	.90	.92	.91

Note. Empty cells indicate no calculation. GFI = goodness-of-fit index; CFI = comparative fit index; TLI = Tucker–Lewis Index; diff = difference.
[a] Difference in the chi-square statistic between a given model and the baseline model. [b] Difference in degrees of freedom between a given model and the baseline model.
** $p < .01$.

Table 4
Correlations With Other Variables

Measure	Sample 1 (n = 182)			Sample 2 (n = 162)			Sample 3 (n = 186)		
	Coefficient α	WFC	FWC	Coefficient α	WFC	FWC	Coefficient α	WFC	FWC
Organizational commitment	.89	−.20*	−.25**						
Job satisfaction	.94	−.36**	−.30**	.93	−.21*	−.16*	.97	−.27**	−.22**
MBI	.82	.56**	.38**	.86	.47**	.19*			
Job tension	.84	.58**	.32**	.82	.43**	.23*	.82	.55**	.38**
Role conflict	.84	.40**	.33**						
Role ambiguity	.85	.39**	.35**						
Intention-to-leave-an-organization	.98	.25**	.23**	.94	.14	.02	.94	.28**	.17*
Search-for-another-job	.88	.12	.18*	.91	.19*	.04	.92	.17*	.19**
Life satisfaction	.87	−.33**	−.44**	.87	−.41**	−.32**	.89	−.53**	−.35**
Relationship satisfaction	.94	−.01	−.16*	.96	−.30**	−.26**	.95	−.27**	−.20**
Relationship agreement	.86	−.14*	−.29**	.87	−.24*	−.20*			
Number of hours worked		.35**	.12		.44**	−.14		.28**	−.13
Sales self-efficacy							.79	−.15*	−.35**
Sales performance							.79	.00	−.38**
Physical symptomology				.92	.34**	.19*	.89	.42**	.37**
ATQ				.95	.29**	.23*	.95	.38**	.40**
Number of children living at home		.21*	.35**		.11	.18*		.07	.19*

Note. WFC = work–family conflict; FWC = family–work conflict. Empty cells indicate that the measure was not included for a given sample. MBI = Maslach Burnout Inventory; ATQ = Automatic Thoughts Questionnaire, a scale used to evaluate depression.
* $p < .05$. ** $p < .01$.

more, the fit indices for this model were again adequate. In summary, some evidence of measurement invariance is apparent across the samples, further establishing the structure of the WFC and FWC scales.

Validity Assessment

Correlations. Table 4 presents the zero-order correlations between WFC and FWC and measures dealing with variables other than WFC and FWC. Negative correlations were predicted between organizational commitment and job satisfaction and WFC and FWC. Across the samples, all of these correlations were significant. We predicted that MBI, job tension, role conflict, role ambiguity, intention-to-leave-an organization, and search-for-another-job would be positively correlated with the WFC and FWC scales. Across the samples, 22 of the 26 correlations pertaining to these predictions were significant. Three of the 4 correlations between WFC and FWC and self-efficacy and sales performance were negative and significant, as predicted.

Life satisfaction, relationship satisfaction, and relationship agreement were predicted to be negatively related to the WFC and FWC scales. Of 16 correlations pertaining to these predictions, 15 were significant across the samples. All positive correlations pertaining to the predictions between WFC and FWC and physical symptomology and ATQ-Depression score were significant, and 4 of the 6 correlations of WFC and FWC with the number of children living at home were significant.

Correlational tests. We also predicted that WFC would be more highly correlated with MBI, job tension, and the number of hours worked than FWC. To test these predictions, we performed t tests between dependent correlations (Cohen & Cohen, 1983, pp. 56–57). The WFC-MBI correlation was compared with the FWC-MBI correlation, the WFC-job tension correlation was compared with the FWC-job tension correlation, and the WFC-number of hours worked correlation was compared with the FWC-number of hours worked correlation.

For Sample 1, WFC was more highly correlated with MBI than was FWC, $t(179) = 2.96$, $p < .01$; WFC was more highly correlated with job tension than was FWC, $t(179) = 4.19$, $p < .01$; and WFC was more highly correlated with the number of hours worked than was FWC, $t(179) = 4.05$, $p < .01$. For Sample 2, WFC was more highly correlated with MBI than was FWC, $t(154) = 4.24$, $p < .01$; WFC was more highly correlated with job tension than was FWC, $t(154) = 2.77$, $p < .01$; and FWC was more highly correlated with the number of hours worked than was FWC, $t(154) = 7.62$, $p < .01$. For Sample 3, WFC was more highly correlated with job tension, $t(182) = 2.64$, $p < .01$, and the number of hours worked, $t(182) = 6.35$, $p < .01$, than was FWC.

For Sample 3, we predicted that FWC would be more highly correlated with self-efficacy and sales performance than WFC. These predictions were supported. FWC was more highly correlated with self-efficacy than was WFC, $t(182) = 3.09$, $p < .01$, and FWC was more highly correlated with sales performance than was WFC, $t(182) = 5.85$, $p < .01$.

Mean-level difference tests between WFC and FWC. To test the prediction that the WFC scale has a higher

mean score than the FWC scale, we summed scores across scale items, and paired t tests between WFC and FWC were performed. For Sample 1, the mean for WFC (M = 15.42, SD = 6.66) was greater than the mean for FWC (M = 9.99, SD = 4.93), $t(178) = 11.33$, $p < .01$. For Sample 2, the mean for WFC (M = 17.16, SD = 8.14) was greater than the mean for FWC (M = 10.30, SD = 5.23), $t(154) = 10.32$, $p < .01$. For Sample 3, the mean for WFC (M = 17.49, SD = 6.94) was again greater than the mean for FWC (M = 11.75, SD = 6.90), $t(182) = 10.20$, $p < .01$.

Discussion

Summary

This article has presented a study designed to develop and validate short, self-report measures of WFC and FWC. To this end, five-item scales of WFC and FWC were developed. The scales showed adequate levels of internal consistency, dimensionality, and discriminant validity across three samples. Also, for numerous on-job and off-job variables, significant correlations with the WFC and FWC scales were found as evidence of construct validity. Several tests assessing differences between correlations supported construct validity, as did mean-level differences tests between WFC and FWC.

As stated previously, existing measures of WFC and FWC have varied widely in terms of reliability and validity, potentially affecting the predictive validity of these constructs. We feel that the measures developed in the present study have some distinct advantages over WFC and FWC measures used in previous research.

First, some studies have used single-item measures of the constructs (Rice et al., 1992; Voydanoff, 1988). It is widely held that single-item measures suffer from random measurement error and may not adequately assess the domain of the construct (Nunnally, 1988; Schriesheim et al., 1993). The measures we have developed are multi-item, exhibit adequate levels of internal consistency, and assess the domain of some commonly agreed on aspects of WFC and FWC.

Second, some studies have operationalized the constructs with long measures. Such measures can be cumbersome for a respondent and do not enhance psychometric properties. For example, 39 items have been used to assess WFC (Burke, 1988; Burke et al., 1979). Although these items sampled the effect of job demands on nine areas of personal, home, and family life, the scores on these items were summed to form an overall composite, ignoring dimensionality. Furthermore, although the reliability of this 39-item measure was high (i.e., α = .92), measures with several items will yield higher coefficient alpha estimates than measures with fewer items, other things being equal. For example, Cortina (1993) and Podsakoff and MacKenzie (1994) noted the importance of taking the number of scale items into consideration when evaluating what constitutes adequate levels of coefficient alpha. They concluded that scales with fewer items are preferable to scales with many items, given comparable coefficient alpha and construct validity estimates. Our five-item scales had coefficient alpha levels ranging from .83 to .89, with an average alpha of .88 for WFC, and of .86 for FWC across the samples. Thus, we feel our scales have the potential for high internal consistency and pose a lesser burden to respondents.

Third, several studies have combined WFC and FWC scales into one measure, ignoring the conceptual distinction between the two constructs (Cooke & Rousseau, 1984; Kopelman et al., 1983; Thomas & Ganster, 1995; Wiley, 1987). As recent literature suggests, WFC and FWC are distinct constructs. The approach taken in the present study acknowledged this conceptual distinction and provided empirical evidence demonstrating discriminant validity between our WFC and FWC scales.

Fourth, studies that have offered separate WFC and FWC measures have used items that reflect potential outcomes of the constructs rather than their content domain, that is, items that assess somatic, physical, and mental symptoms such as being unable to sleep because of WFC, FWC, or both (Bedeian et al., 1988; O'Driscoll et al., 1992; Parasuraman et al., 1989; e.g., "I can't sleep because of thinking about things at work that I have to get done"; O'Driscoll et al., 1992, p. 279). We feel that the content domains of our WFC and FWC scales reflect only aspects of work interfering with performing family-related duties and vice versa.

Fifth, other studies have used measures that have not been subjected to rigorous scale development (Frone et al., 1992; Gutek et al., 1991; Judge et al., 1994). Although these measures do seem to possess adequate content validity and internal consistency, they have not been scrutinized as rigorously with respect to construct validity as have our WFC and FWC measures. Furthermore, the coefficient alpha estimates of these other WFC and FWC measures were generally lower than the coefficient alpha estimates of our measures. For example, Gutek et al. (1991) reported alpha estimates of .81 and .83, and .79 and .83 for four-item measures of WFC and FWC: Judge et al. (1994) reported alpha estimates of .82 and .76 for four-item measures of WFC and FWC. As stated above, we report an average coefficient alpha of .88 for WFC, and of .86 for FWC.

Finally, we examined several studies that used some type of multiple-item measures of WFC and FWC as separate constructs (Bedeian et al., 1988; Cooke & Rousseau, 1984; Frone et al., 1992; Gutek et al., 1991; Judge et al., 1994; O'Driscoll et al., 1992; Parasuraman et al., 1989; Wiley, 1987) and compared the average correlations they reported to those we report. Our WFC and

FWC measures consistently showed stronger correlations with job satisfaction, organizational commitment, job tension, and life satisfaction in our study than did the corresponding measures of WFC and FWC used in the previously described studies. These stronger correlations, we feel, added strength to the potential predictive validity of our scales.

Limitations and Future Research

The study presented here is not without limitations. First and foremost, the scales derived in this study are not as useful as scales that use a multidimensional approach to the measurement of WFC and FWC. Although our scales assess a general demand and time- and strain-based conceptualization, some researchers advocate multi-item scales assessing separate dimensions of general demand and time- and strain-based WFC and FWC (e.g., Greenhaus & Beutell, 1985). This approach could provide valuable insight into how separate WFC and FWC aspects relate to on- and off-job attitudes and behaviors. Thus, future studies may want to attempt to expand the measurement of WFC and FWC with a multidimensional approach.

Second, all measures relating to variables other than WFC and FWC in our study were of a self-report nature, and our study was nonexperimental in design. Because only experiments can offer evidence of causality, all that can be concluded from our study is that the WFC and FWC scales were related to on- and off-job constructs at one point in time.

Finally, though three different samples were represented in the present article, validation of the scales across numerous occupations is needed. It is hoped that further validations will lend confidence to the use of the scales, as well as add to the generalizability of WFC and FWC research.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Bachrach, S. B., Bamberger, P., & Conley, S. (1991). Work–home conflict among nurses and engineers: Mediating the impact of role stress on burnout and satisfaction at work. *Journal of Organizational Behavior, 12,* 39–53.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16,* 74–94.

Bandura, A. (1986). *Social foundations of thought and action: A social cognition theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bedeian, A. G., Burke, B. G., & Moffett, R. G. (1988). Outcomes of work–family conflict among married male and female professionals. *Journal of Management, 14,* 475–491.

Behrman, D., & Perreault, W. (1984). A role-stress model of the performance and satisfaction of industrial salespersons. *Journal of Marketing, 48,* 9–21.

Bentler, P. M. (1990). Comparative fit indices in structural equation modeling. *Psychological Bulletin, 107,* 238–246.

Bluedorn, A. C. (1982). A unified model of turnover from organizations. *Human Relations, 35,* 135–153.

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

Brown, S. P., & Peterson, R. A. (1994). The effect of effort on sales performance and job satisfaction. *Journal of Marketing, 58,* 70–80.

Burke, R. J. (1988). Some antecedents and consequences of work–family conflict. *Journal of Social Behavior and Personality, 3,* 287–302.

Burke, R. J., Weir, T., & Duwors, R. E. (1979). Type A behavior of administrators and wives' reports of marital satisfaction and well-being. *Journal of Applied Psychology, 64,* 57–65.

Cascio, W. (1991). *Costing human resource: The financial impact of behavioral organizations.* New York: Wiley.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlational analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooke, R. A., & Rousseau, D. M. (1984). Stress and strain from family roles and work-role expectations. *Journal of Applied Psychology, 69,* 251–262.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104.

DeVillis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18,* 39–50.

Frone, M. R., Russell, M., & Cooper, M. L. (1992). Antecedents and outcomes of work–family conflict: Testing a model of the work–family interface. *Journal of Applied Psychology, 77,* 65–78.

Greenhaus, J. H. (1988). The intersection of work–family roles: Individual, interpersonal, and organizational issues. *Journal of Social Behavior and Personality, 3,* 23–44.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review, 10,* 76–88.

Gutek, B. A., Repetti, R., & Silver, D. (1988). Nonwork roles and stress at work. In C. Cooper & R. Payne (Eds.), *Causes, coping, and consequences of stress at work* (pp. 147–174). New York: Wiley.

Gutek, B. A., Searle, S., & Klepa, L. (1991). Rational versus gender role expectations for work–family conflict. *Journal of Applied Psychology, 76,* 560–568.

Hendrix, W. H., Nestor, K. O., & Troxler, R. G. (1985). Behavioral and psychological consequences of stress and its antecedent factors. *Journal of Applied Psychology, 70,* 188–201.

Hollon, S. D., & Kendall, P. C. (1980). Cognitive self-statements in depression: Development of an automatic thoughts questionnaire. *Cognitive Therapy and Research, 4,* 383–395.

House, R. J., & Rizzo, J. R. (1972). Role conflict and ambiguity as critical variables in a model of organizational behavior. *Organizational Behavior and Human Performance, 7,* 467–505.

Hoyle, R. H., & Panter, A. (1994). Writing about structural

equation models. In R. H. Hoyle (Ed.), *Structural equation modeling concepts, issues, and applications* (pp. 158–176). Newbury Park, CA: Sage.

Jones, A. P., Johnson, L. E., Butler, M. C., & Main, D. S. (1983). Apples and oranges: An empirical comparison of commonly used indices of interrater agreement. *Academy of Management Journal, 26,* 507–519.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL VII: A guide to program and applications.* Chicago: SPSS.

Judge, T. A., Boudreau, J. W., & Bretz, R. D. (1994). Job and life attitudes of male executives. *Journal of Applied Psychology, 79,* 767–782.

Kahn, R. L. (1981). *Work and health.* New York: Wiley.

Kahn, R. L., & Byosiere, P. (1992). Stress in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 571–650). Palo Alto, CA: Consulting Psychologists Press.

Kahn, R. L., Wolfe, D. M., Quinn, R., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress.* New York: Wiley.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations.* New York: Wiley.

Kemery, E. R., Mossholder, K. W., & Bedeian, A. G. (1987). Role stress, physical symptomology, and turnover intentions: A causal analysis of three alternative specifications. *Journal of Occupational Behavior, 8,* 11–23.

Kopelman, R. E., Greenhaus, J. H., & Connolly, T. F. (1983). A model of work, family, and interrole conflict: A construct validation study. *Organizational Behavior & Human Performance, 32,* 198–215.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and predictions tests: Their reliability and validity. *Marriage and Family Living, 21,* 251–255.

Marsh, H. W. (1995). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1,* 5–34.

Marsh, H. W., & Hocevar, D. P. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. *Psychological Bulletin, 97,* 562–582.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2,* 99–113.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior, 14,* 224–247.

Nunnally, J. (1988). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Driscoll, M. P., Ilgen, D. R., & Hildreth, K. (1992). Time devoted to job and off-job activities, interrole conflict, and affective experiences. *Journal of Applied Psychology, 77,* 272–279.

Parasuraman, S., Greenhaus, J. H., Rabinowitz, S., Bedeian, A. G., & Mossholder, K. H. (1989). Work and family variables as mediators of the relationship between wives' employment and husbands' well-being. *Academy of Management Journal, 32,* 185–201.

Pennebaker, J. W. (1982). *The psychology of physical symptoms.* New York: Springer-Verlag.

Pleck, J. H., Staines, G. L., & Lang, L. (1980). Conflicts between work and family life. *Monthly Labor Review, 103,* 29–32.

Podsakoff, P. M., & MacKenzie, S. B. (1994). An examination of the psychometric properties and nomological validity of some revised and reduced substitutes for leadership scales. *Journal of Applied Psychology, 79,* 702–713.

Price, J. P., & Mueller, C. W. (1986). *Handbook of organizational measurement.* Marshfield, MA: Pittman.

Quinn, R. P., & Staines, G. L. (1979). *The 1977 quality of employment survey.* Ann Arbor, MI: Institute for Social Research, University of Michigan.

Rice, R. W., Frone, M. R., & McFarlin, D. B. (1992). Work–nonwork conflict and the perceived quality of life. *Journal of Organizational Behavior, 13,* 155–168.

Rizzo, J., House, R. J., & Lirtzman, S. I. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly, 15,* 150–163.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 1–15). San Diego, CA: Academic Press.

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management, 19,* 384–417.

Staines, G. L., & Pleck, J. H. (1984). Nonstandard work schedules and family life. *Journal of Applied Psychology, 69,* 515–523.

Sujan, H., Weitz, B. A., & Kumar, N. (1994). Learning orientation, working smart, and effective selling. *Journal of Marketing, 58,* 39–52.

Thomas, L. T., & Ganster, D. C. (1995). Impact of family-supportive work variables on work–family conflict and strain: A control perspective. *Journal of Applied Psychology, 80,* 6–15.

Voydanoff, P. (1988). Work role characteristics, family structure demands, and work–family conflict. *Journal of Marriage and the Family, 50,* 749–761.

Wiley, D. L. (1987). The relationship between work–nonwork role conflict and job-related outcomes: Some unanticipated findings. *Journal of Management, 13,* 467–472.

(Appendix follows on next page)

Appendix

Items on the Scales

Work–Family Conflict Scale

1. The demands of my work interfere with my home and family life.
2. The amount of time my job takes up makes it difficult to fulfill family responsibilities.
3. Things I want to do at home do not get done because of the demands my job puts on me.
4. My job produces strain that makes it difficult to fulfill family duties.
5. Due to work-related duties, I have to make changes to my plans for family activities.

Family–Work Conflict Scale

1. The demands of my family or spouse/partner interfere with work-related activities.
2. I have to put off doing things at work because of demands on my time at home.
3. Things I want to do at work don't get done because of the demands of my family or spouse/partner.
4. My home life interferes with my responsibilities at work such as getting to work on time, accomplishing daily tasks, and working overtime.
5. Family-related strain interferes with my ability to perform job-related duties.

Received April 17, 1995
Revision received February 29, 1996
Accepted March 2, 1996 ■