For information address:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Structural equation modeling: concepts, issues, and applications / edited by Rick H. Hoyle
p. cm.
Includes bibliographical references and index.
ISBN 0-8039-5317-8 (acid-free). — ISBN 0-8039-5318-6 (pbk.: acid-free)
1. Social sciences—Mathematical models. 2. Social sciences—Statistical methods. I. Hoyle, Rick H.
H61.25.S767 1995
300'.1'5118—dc20 94-47262

This book is printed on acid-free paper.

97 98 99 00 01 10 9 8 7 6 5 4

Sage Project Editor: Susan McElroy

Brief Contents

Contents

Fiske guidelines, CFA models, or CDP models begin with single indicators of each trait/method combination even though each trait/method combination is often based on multiple indicators (e.g., the individual items composing a scale). There are, however, important advantages in incorporating the multiple indicators into models like those considered here (see Marsh, 1993; Marsh & Hocevar, 1988). Although MTMM studies typically consider data from only one group, Marsh et al. (1993) demonstrated pragmatic and substantive benefits in testing the invariance of solutions across multiple groups. There may also be advantages in testing invariance constraints within a single group for CFA (Marsh et al., 1993) or direct-product (Bagozzi & Yi, 1992) models. MTMM studies typically focus on a "within-construct" component of construct validation, but Marsh (1988) and Marsh et al. (1993) demonstrated how external validity criteria can be incorporated into CFA-MTMM models.

In conclusion, this chapter has an important message for applied researchers who wish to use the MTMM paradigm. MTMM data have an inherently complicated structure that will not be fully described in all cases by any of the models or approaches typically considered. There is, apparently, no "right" way to analyze MTMM data that works in all situations. Instead, we recommend that researchers consider several alternative approaches to evaluating MTMM data—an initial inspection of the MTMM matrix using the Campbell-Fiske guidelines followed by fitting at least the subset of CFA models recommended by Marsh and Grayson (in press) and the CDP model. The Campbell-Fiske guidelines should be used primarily for formative purposes, the CDP model seems most appropriate for summative evaluations of the extent to which the MTMM data fulfill the Campbell-Fiske guidelines, and the CFA models apparently serve both summative and formative purposes. It is, however, important that researchers understand the strengths and weaknesses of the different approaches. For each of the different latent-variable approaches, researchers should evaluate results in relation to technical considerations such as convergence to proper solutions and goodness of fit, but they should also place more emphasis on substantive interpretations and theoretical considerations. Despite the inherent complexity of MTMM data, we feel confident that the combination of common sense, a stronger theoretical emphasis on the design of MTMM studies, a better quality of measurement at the level of trait-method units, an appropriate arsenal of analytical tools such as recommended here, and a growing understanding of these analytic tools will allow researchers to use the MTMM paradigm effectively.

11 Sex-Race Differences in Social Support and Depression in Older Low-Income Adults

JANE A. SCOTT-LENNOX

RICHARD D. LENNOX

Biomedical research, after years of using a universal (young, white male) model, has been forced to recognize the important influences of sex, race, age, and other characteristics on disease processes and treatment outcomes. Social science has long recognized sex, race, and their interaction as fundamental stratifying attributes in American society. They help define social roles and responsibilities, resources, socialization, and experiences a person has throughout life. Lifelong exposure to health hazards, economic disadvantage, and limited access to health care place female and black American rural elders at greater risk of physical disability, multiple chronic illnesses, and poverty (Crystal & Shea, 1990). Chronic stresses such as these increase the risk of depressive symptoms, which, in turn, reduce the effectiveness of medical treatments and discourage interaction with others, thereby perpetuating

AUTHORS' NOTE: We would like to thank Jane E. Morrow, E. Michael Bohlig, Lore K. Wright, Deborah L. Gold, Rick Hoyle, and an anonymous reviewer for helpful comments on earlier drafts of this chapter, and Dr. Harold L. Cook for the use of the data analyzed here.

The research was supported by NIH Multipurpose Arthritis Center Grant AM-30701 to the University of North Carolina at Chapel Hill, by the Numerical Science Core of the UNC Thurstone Multipurpose Arthritis Center, and by a National Institute of Aging Post-Doctoral Research Training Fellowship grant to the Center for the Study of Aging and Human Development at Duke University Medical Center.

the chronic stress cycle (Blazer, 1982). Despite these disadvantages, or perhaps because of them, high levels of informal social support—the care and help received from family members and friends—often are observed among populations at the most risk of chronic stress (Mercier & Powers, 1984).

As its label implies, social support theory generally assumes the universal utility of social relationships: If available and needed, supportive relationships with other people should help protect an individual from stress and its consequences. Hence, high levels of available support among disadvantaged groups would be expected to reduce their experience of stress and prevent distress. However, supportive relationships are embedded in social networks consisting of bivalenced ties—social relationships that contribute to stress in people's lives, even though they are supportive on other dimensions. The amount of support needed, the size of the network available to provide it, and even the characteristics of supporters all contribute to the adequacy of a social support network and to the extent of dependence on bivalenced networks.

Studies that find that the types and amounts of support available vary depending on a person's sex and/or race do not violate this basic assumption of the universal utility of support (e.g., Chatters, Taylor, & Jackson, 1985; Palinkas, Wingard, & Barrett-Connor, 1990; Scott & Roberto, 1985). However, studies also have found that the same social support resources differ in their ability to reduce or prevent distress for men and women (Dean, Kolody, & Wood, 1990; Haines & Hurlbert, 1992), for blacks and whites (Krause, 1989; Quevillon & Trennery, 1983; Vaux, 1985), or for men and women of the same race (Husaini et al., 1991). The source of supportive relationships and the extent to which the support source affects the proportion of bivalenced ties in the social network are important qualifiers to gender differences (Leffler, Krannick, & Gillespie, 1986), and perhaps race differences, in the benefits of social support.

In one of the most intriguing studies of gender differences in social support, Haines and Hurlbert (1992) found that three dimensions of support network range (density, diversity, and size) differentially affect exposure to stress for men and women, access to social support, and level of distress. Because their analyses were based on data from a population survey that could not fully control for major life stress associated with poverty or chronic illness, the complex relationships among support, stress, and depression implied by chronic stress theories are only partially explicated by these results. In addition, their analyses did not explore moderating effects of race or of sex-race interactions suggested by Husaini et al. (1990, 1991) and others.

Universal Versus Group-Specific Models of Support and Depression

The coincidence of high levels of social resources on one hand and high exposure to chronic stressors on the other, when combined with the many stratifying effects of sex and race, raises the following question: Given comparable levels of chronic stress and similar life circumstances, are the same social support resources equally available and valuable for reducing psychological distress, regardless of moderating effects of sex, race, or their interaction (sex-race)? Empirically, the most precise answer to this question lies in results of multigroup structural equation modeling (MSEM), wherein dimensions of support are used to predict depression for sex-race subgroups. When the theory underlying the model indicates that a mediating relationship among predictors of outcomes may vary by population subgroups, MSEM is preferable to multiple linear regression path-analytic models.

MSEM is uniquely suited for exploring this question because of its ability to test a theoretical model for its applicability to different groups simultaneously. MSEM models do not require cumbersome interaction terms and nested models to estimate hypothesized group differences in path-analytic model coefficients or model fit. A single χ^2 goodness-of-fit statistic evaluates a set of complex models—one for each group. To validate the usual assumption that groups are equivalent, groups can be required to have identical estimates for all parameters (a "fully constrained" or universal model). Differences among groups can be evaluated for their appropriateness by "freeing" some parameters (allowing one or more groups to vary uniquely), "fixing" (setting parameters to zero), and/or "constraining" (requiring two or more groups to have equal parameters) any or all parameters for different groups. MSEM analyses often begin by estimating a fully constrained model (equivalent to a full sample regression model), then relaxing constraints to allow for group-specific differences in particular parameters based on theory or inductive evidence (e.g., Lagrange multiplier tests for impact of freeing one or more constraints on model fit).

In deference to the substantial literature on social support and depression in older adults (for reviews, see Blazer, 1982; George, 1989),

our study focuses on two dimensions of social networks that are likely to impact the mental health of older adults undergoing chronic stress. The first and most powerful social support dimension is a person's perception of the adequacy of her or his support network. Because evaluation of the adequacy of support resources is a function of resources *desired*, *available*, and *used* (Krause, 1989), satisfaction with support is a psychological experience in its own right and thus is closely associated with depressive symptoms and other measures of mental health. In contrast, quantitative or structural aspects of a person's social support network help determine whether a person is satisfied with her or his support, and only indirectly contribute to depressive illness. From the path-analytic perspective, this places satisfaction with support as a mediator in the linear relationship between support network characteristics and depression (Lin, 1986).

Based on Haines and Hurlbert (1992), three measures of network range are expected to predict satisfaction with support: network size, network density, and sex composition of network. People with larger social networks (size) are assumed to have, on average, access to more social resources overall, and thus a greater probability of having access to appropriate and desired social support. If so, network size should be an important predictor of satisfaction with available support and of mental health. Social resources also are more likely to be rallied to the person's support if the members of the network know one another well (high network density). Finally, differential gender roles that occupy today's social environment may translate into the provision of different types and amounts of social support. All other things equal, this would suggest that people whose networks contain both men and women are more likely to have access to a wider range of support than are those whose networks consist only of same-sex network members. When considering self-reported mental health, however, these structural features of social networks are likely to be useful only if people perceive them to be adequate and desirable. Therefore, objective components of network size, network density, and sex homophily are expected to affect self-reported mental health only as mediated through perceived adequacy of social support.

This theoretical model is illustrated in the path diagram in Figure 11.1, wherein three intercorrelated network structure components (size, density, and sex homophily) determine perceived support adequacy, which in turn predicts an individual's level of depressive symptoms. The three network structure variables do not directly influence depression in this model. Of course, all things are not equal, particularly when considering the various socioeconomic roles and resources allocated by gender and race. For this reason, we suspect that the adequacy of this model, or at least specific parameters in it, is likely to differ for men and women of different races.

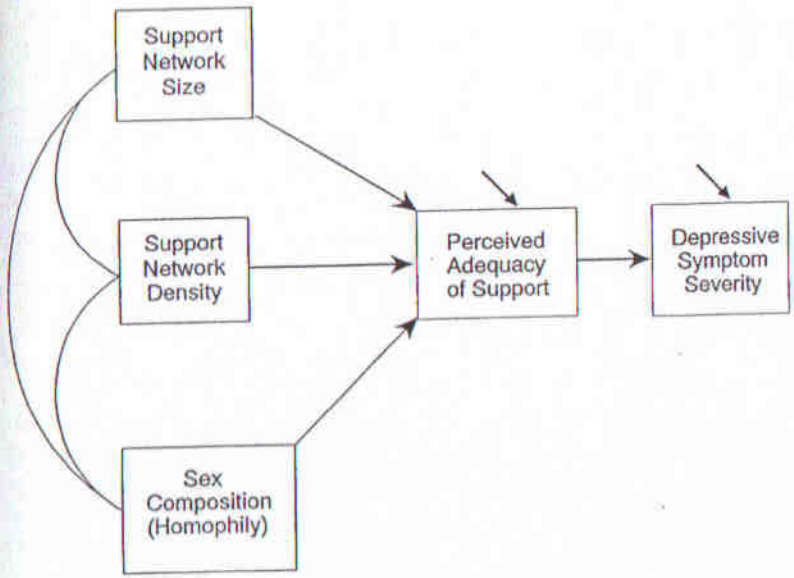

Figure 11.1. A Model of the Relationship Among Social Support Network Structure, Perceived Adequacy of Social Support, and Depressive Symptoms

TESTING THE EMPIRICAL MODEL

Three multigroup structural equation models are evaluated to test the theorized mediating effect of satisfaction with support, illustrated in Figure 11.1, as well as the moderating effects of sex and race. Comparison of the goodness-of-fit indexes for the models allows us to evaluate the appropriateness of each model.

For each group, depressive symptom severity is the ultimate dependent variable. It is predicted by perceived adequacy of social support, which mediates the effect of network size, density, and sex homophily. Residual variances in the dependent variables (depressive symptom

severity and perceived adequacy of social support) not explained by the model are estimated as well. Network size, density, and sex homophily are free to correlate. Because all effects of support networks are expected to be mediated through perceived adequacy of support, their direct effects on depression are assumed to be zero, except in the final model, where the appropriateness of this assumption is formally tested.

The first model tested is a "universal" model that constrains all sex-race subgroups to have equal parameter estimates. It is comparable to a traditional path analysis for the full sample. Sex and race differences in the utility of support resources detected in past studies are assumed to be artifacts of the higher incidence of disadvantage experienced by some groups compared to others. Presumably, neither sex nor race is responsible for observed differences in depression but, rather, the higher incidence of disability, poverty, illness, and levels of support available that covary with sex and race is responsible. Once income, disability levels, and rural residence are controlled (as was done through the sampling frame in these data), the universal model hypothesizes (a) no significant group differences in parameter estimates and (b) equivalent model fit for all groups.

The second, "group-sensitive," model estimated acknowledges that sex-race subgroups may have distinct relationships among support and depression, despite the general applicability of the theoretical model. Its theoretical basis arises from Haines and Hurlbert (1992) and other studies that have found that sex or race stratify the utility of support resources for promoting psychological well-being. Extending these single modifier studies, our model assumes that the interaction of sex and race may predict different relationships in the theoretical model of depression and support. To identify significant group-specific differences, Lagrange multiplier (LM) tests of equality constraints across samples estimated in the universal model were examined. Equality constraints were removed only if they dramatically improved model fit. A change in χ^2 of 5.0 or more points per degree of freedom (i.e., $p < .01$) was considered significant, rather than the usual 3.84 (when df = 1) or less.

Finally, in addition to freeing the parameters identified in the group-sensitive model, a third "direct-effect" model tested the assumption that most of the effects of support structure on depressive symptoms are mediated through satisfaction with support.

Method

SAMPLE

Assessing the effects of informal support among economically disadvantaged Americans precludes the use of patients who present themselves for treatment. Barriers to treatment associated with race, age, and economic disadvantage create a selection bias that would adversely affect the validity of the study. Therefore, we used data from a survey designed specifically to sample multiply disadvantaged rural elders who had arthritis (Cook et al., 1986). To participate in the study, a person had to (a) reside in a rural region of a predominantly rural Southern county, (b) have an annual household income at or below 185% of the federally established poverty level, and (c) believe he or she had arthritis. Because there were no rheumatology clinics in the county, and because low-income rural residents are less likely to use specialty medical care, subjects were recruited from lists of names referred by local general practice physicians, social service and county extension agents, enrolled study participants, and indigenous lay interviewers. Telephone screening and scheduling excluded the poorest candidates, whose homes had no telephone. Of those that could be contacted, roughly equal numbers of black and white Americans of each sex were used to test for the interaction of sex and race.

Demographics. Of the 219 subjects interviewed, 214 had complete data on all variables. Normal age distributions associated with arthritis generated a predominantly older sample with more than two thirds of the subjects age 60 or older (range = 32 to 88 years); however, mean age did not differ significantly across subgroups, $F(3, 210) = 1.59$, $p =$ ns. Years of education, $F(3, 210) = 5.04$, $p < .01$; percent married, $F(3, 210) = 12.38$, $p > .01$; percent widowed, $F(3, 210) = 19.54$, $p > .01$; household size, $F(3, 210) = 4.44$, $p > .01$; and proportion living alone, $F(3, 210) = 5.77$, $p > .01$, varied by sex, race, or their interaction. Women of both races completed more years of formal education than did black males and were more likely to live alone, with almost half (48%) of the white females living alone. Equally as many black males as white females were currently married, but black males reported substantially larger average household size than white females (2.4 vs. 1.2, respectively). Black females were dramatically more likely to be

widowed than any other group, but many black widowed women still lived with others.

Health Status. Sex-race subgroups had comparable levels of functional limitations, pain severity, and years with arthritis, but white respondents reported more doctor diagnosed chronic illnesses, $F(3, 210) = 5.75$, $p < .01$, possibly a reflection of greater lifetime use of medical services than a true difference in health status across races. Epidemiologic studies would not have anticipated differences in type of arthritis observed in this sample (Cunningham & Kelsey, 1984). In particular, the overrepresentation of white males with rheumatoid arthritis (37.9%, compared with 20.4% for white females, 13.2% for black males, and 16.7% for black females) indicates a marked difference in health status among groups. Absence of differences in levels of functional disability or pain severity notwithstanding, there may be psychological or social consequences of having rheumatoid arthritis versus other type of arthritis that may have contributed to differences observed here.

PROCEDURE

In-home, 1.5- to 2-hour structured interviews conducted by sex- and race-matched interviewers provided self-report information on subjects' health status, functional limitations, social networks, and depressive symptoms, and authorization for rheumatologists to examine subjects' health records. For subjects whose medical records did not report the results of a standard rheumatological examination, arthritis diagnosis and physician assessment of functional limitations were obtained during examinations conducted by licensed rheumatologists associated with the project. Although 12.5% of the respondents were not examined by a rheumatologist, arthritis diagnosis was confirmed in 95% of all subjects who underwent a diagnostic examination, suggesting a high correspondence between self- and physician-diagnosis of arthritis in this sample.

MEASURES

Depressive Symptoms. The dependent variable, depressive symptom severity, was operationalized as the number and severity of self-reported depressive symptoms experienced during the two weeks preceding the interview using items from the Center for Epidemiological Studies Depression Scale (Radloff, 1977). The summary score used here is based on a total of the 16 negatively worded symptoms to avoid educational confounding of positively worded symptoms reported by Lin (1989). Each item has a possible range of 0 to 3 (0 = *not at all* to 3 = *5 or more days in the last two weeks*); hence, the summary score had a possible range from 0 to 48; 43 was the highest reported score in these data. Internal consistency of the unit-weighted CESD16 scale was high (coefficient alpha = .90).

Perceived Adequacy of Social Support (PASS). How satisfied a person was with her or his network was operationalized as the sum of responses to questions adapted from the Instrumental-Expressive Support Scale (Ensel & Woelfel, 1986) and from measures developed by Strogatz (Strogatz & James, 1986) to detect desire for more support. Scores for how often a respondent was bothered by not having a close companion, not seeing enough of people he or she feels close to, not having enough close friends, and not having someone who shows love and affection (with values ranging from 1 = *all the time* to 5 = *never*) were added to scores for whether the respondent wished she or he knew more people to talk with, to have a "good time" with, to depend on for help with chores or errands, and to rely on for help with arthritis-related problems (coded 1 = *wish I knew more people* to 5 = *already know enough people*). The unit-weighted sum of these eight items appeared to be internally consistent (coefficient alpha = .80) and resulted in the Perceived Adequacy of Support Scale (PASS), which ranged from extremely dissatisfied (coded as 8) to completely satisfied (coded as 40) with the quality and quantity of a subject's social support network.

Measures of Social Support Networks. To identify who subjects considered to be members of their social networks, the survey included two network "name generator" questions: one identifying all the people close to the respondent (a) who had (or would have) helped during a hard time or emergency and (b) with whom the respondent discussed important matters. *Network size* was operationalized as a count of all names elicited in response to the two network name generators (observed range = 0 to 13). *Network density* and *sex homophily* measured the structure and sex composition of a respondent's network. Density, based on "core" networks (the first five names mentioned in response to either question; Marsden, 1987), is a measure of how close the

respondent believes network members are to each other. It is the mean of tie-level (member 1's relationship to member 2 and so on) variables that were coded 0 if the network members were strangers to one another, 1 if they were very close, and .5 if they were neither close nor strangers (range: 0 = *completely disconnected* to 1 = *completely connected*). Sex homophily, the proportion of the respondents' total support network members that were the same sex as the respondent, was created by coding each network member as same sex (1) or opposite sex (0) of respondent, the sum of which was then divided by total network size (range: 0 = *all opposite sex* or *no network* to 1 = *all same sex*).

STATISTICAL ANALYSES

Mean Comparisons. Group differences in access to support were evaluated as mean differences in each of the support measures across subgroups. Having controlled for socioeconomic and health status variations usually argued to determine mean psychological well-being and social support, mean support levels across sex-race subgroups were compared using analysis of variance.

Multigroup Structural Equation Models. Covariance matrices of the variables for each group were analyzed using EQS (Bentler, 1992a). In the first model, all parameters were constrained to be equal for all groups.

Results

SEX-RACE DIFFERENCES IN ACCESS TO SUPPORT

Table 11.1 presents means, standard deviations, and Pearson product moment correlations for each study variable by sex-race subgroup.

Depressive Symptoms (CESD16). Considering the multiple chronic disadvantages that study respondents were experiencing, it is not surprising that subgroups reported uniformly high levels of depressive symptoms, $F(3, 210) = 1.24$, $p > .29$. Indeed, means for all subgroups exceeded national averages for the 20-item standard scale, which average below 10 out of 60 possible points (Radloff, 1977). However, they

TABLE 11.1 Variable Means and Correlations by Sex-Race Subgroup

Variable	*Mean*	*SD*	*CESD16*	*PASS*	*Size*	*Density*
White Males						
CESD16	11.9	8.4				
PASS	33.9	5.9	−0.26			
Size	2.9	2.0	−0.01	0.01		
Density	0.8	0.4	0.04	−0.15	0.36	
Homophily	0.3	0.3	0.02	−0.00	0.37	0.52
White Females						
CESD16	15.4	10.2				
PASS	30.7	8.3	−0.43			
Size	3.2	1.5	−0.01	0.12		
Density	0.8	0.3	0.17	0.19	0.39	
Homophily	0.5	0.3	0.24	−0.31	0.29	0.30
Black Males						
CESD16	14.9	12.9				
PASS	30.2	8.1	−0.25			
Size	3.8	2.3	−0.11	0.06		
Density	0.8	0.3	−0.11	0.07	0.45	
Homophily	0.4	0.3	0.04	−0.07	0.37	0.02
Black Females						
CESD16	13.7	9.8				
PASS	31.4	7.4	−0.42			
Size	4.2	2.3	0.32	−0.02		
Density	0.8	0.4	0.06	0.04	0.45	
Homophily	0.6	0.3	0.03	0.34	0.32	0.32

NOTE: CESD16 = 16-item Center for Epidemiological Studies Depression Scale; PASS = Perceived Adequacy of Social Support Scale.

were comparable to means reported in samples of disabled adults (Turner & Wood, 1985).

Satisfaction With Support (PASS). Most subjects were satisfied with the number of supporters in their social networks and felt the support they had was adequate. White males reported higher levels of satisfaction with their support networks, on average, but only marginal subgroup differences in means were observed, $F(3, 210) = 2.62$, $p = .052$.

Structure of Support Networks. For all sex-race subgroups except white males, average network size was greater than the 3.01 ties reported in the national probability survey (1985 General Social Survey; GSS) that used a similar name generator (Marsden, 1987). Bonferroni tests found the most significant differences between groups were those between the large networks of black females and small networks of white males, $F(3, 210) = 4.46$, $p < .005$. Networks reported here generally were more dense than those observed in national samples (Marsden, 1987), presumably because of the combined effect of rural residence, poverty, and advanced age, which have each been argued to constrain the range of social networks (Linn, Husaini, Whitten-Stovall, & Broomes, 1989). There were no sex-race differences in mean density, $F(3, 210) = .25$, $p > .86$. A comparison of sex-race subgroup means revealed dramatic differences in sex composition of networks, with same-sex ties more common for women, especially black women, $F(3, 210) = 11.47$, $p < .001$.

THE UNIVERSAL MSEM MODEL

Derivations of Degrees of Freedom. The universal model estimated the theoretical model depicted in Figure 11.1, assuming that sex-race groups would not differ in parameter estimates or model fit. The universal structural equation model is based on four covariance matrices, each with five measured variables, for a total across the samples of 20 variances and 40 covariances (df = 60). For all three models, variances are free to differ by group for all variables. In the universal model, 27 parameters were estimated for all four groups simultaneously: three coefficients for the covariances among the exogenous variables, another three for the direct effects of network variables on perceived social support, one for direct effects of perceived social support on depression, plus the 20 variances (five variables × four groups). Thus the χ^2 for the universal model is based on 33 degrees of freedom (60 possible parameters minus 27 estimated parameters). The group-specific model consumes an additional 5 degrees of freedom to estimate differences implied by LM tests (df = 28). Finally, the model evaluating the appropriateness of the mediation model uses another 12 degrees of freedom (df = 16).

Table 11.2 presents standardized coefficients and goodness-of-fit tests estimated in the fully constrained universal model and in the group-sensitive multisample structural equation models tested here.

TABLE 11.2 Standardized Coefficients for Two Models of Social Network Characteristics (Size, Density, and Sex Homophily), Perceived Adequacy of Social Support (PASS), and Depressive Symptom Severity (CESD16)

	Universal Model Estimates				*Group-Sensitive Model Estimates*			
Parameters	*White Males*	*White Females*	*Black Males*	*Black Females*	*White Males*	*White Females*	*Black Males*	*Black Females*
CESD16-PASS	−.33*	−.39*	−.29*	−.38*	−.32*	−.39*	−.30*	−.38*
PASS-Size	.03	.02	.03	.03	.04	.02	.03	.03
PASS-Density	.00	.00	.00	.00	−.11	.33*	−.07	−.08
PASS-Homophily	−.00	−.00	−.00	−.00	.04	−.44*	−.07	.36*
Size-Density	.39*	.47*	.37*	.37*	.40*	.48*	.41*	.38*
Size-Homophily	.37*	.41*	.26*	.27*	.38*	.42*	.30*	.27*
Density-Homophily	.36*	.33*	.29*	.31*	.45*	.41*	−.04	.37*
CESD16 Equation Error	.94	.92	.95	.93	.95	.92	.95	.93
PASS Equation Error	1.00	1.00	1.00	1.00	1.00	.90	1.00	.94
Goodness-of-Fit Indexes								
Model chi-square	$\chi^2 = 51.94$, df = 33, $p = .02$				$\chi^2 = 24.60$, df = 28, $p = .66$			
Nonnormed fit index	.79				1.04			
Comparative fit index	.83				1.00			

Asterisks indicate parameters that are different from 0 at the .05 level of statistical significance. Group differences in standardized parameter estimates reflect group-specific differences in variances for the variable.

Not surprisingly, the fully constrained model produced a less than desirable fit to the observed data. The nonnormed fit index (= .79) was unimpressive, as was the comparative fit index (= .83). From an inferential perspective, the fit between the observed and modeled covariance matrix was reasonably good, $\chi^2(33, N = 214) = 51.94$, $p = .02$, although it was statistically significant at less than the .05 level, which indicates the fit may be improved by freeing some of the equality constraints.

THE GROUP-SENSITIVE MODEL

LM tests indicated that the removal of five equality constraints would significantly improve the fit of the model. The LM tests indicated that releasing the requirement that network density and sex homophily be associated at the same level for the white and black males would improve the fit of the model, $\Delta\chi^2(1, N = 214) = 8.06$, $p = .005$. Evidence of this inequality is reflected in the correlations between sex homophily

and network density (see Table 11.1) for these two groups. The r for density-homophily was .51 for white males but only .02 for black males; correlations for women of both races were nearly identical, although somewhat lower than for white males. These differences suggest that the tendency for dense networks to comprise members of the same sex does not hold for black males but does for all other subgroups, particularly white males.

Allowing group-specific estimates for the relationship between sex homophily and perceived social support also should significantly improve the fit of the model. Specifically, men of both races appeared to be more similar in the effects of sex ratio on satisfaction with support than were women; however, white and black women had opposite relationships between satisfaction with support and the sex composition of their networks. This is reflected in the correlation (Table 11.1) between PASS and sex homophily; for white women, the correlation is substantial and negative ($r = -.31$), whereas black women have a strong positive association between same-sex networks and satisfaction with support ($r = .34$). The LM test indicated that allowing groups to have unique estimates of the degree of association between sex homophily and satisfaction would reduce the χ^2 by at least 10.66 ($df = 3, p < .005$).

Finally, race and sex also interacted to predict the relationship between network density and perceived social support. Dense networks were associated with low perceived social support in the white male sample but with high perceived social support among white females; however, network density was not a strong predictor of satisfaction with support for either black men or women in this sample. LM test results indicated that releasing the constraint that white females be equal to other groups would reduce the χ^2 by 5.46 ($df = 1, p = .019$).

The appropriateness of the implied group-specific model was confirmed by the expected impact both on parameter estimates and on model fit. The following constraints were removed to capture significant group-specific differences in the universal model:

1. White females were allowed a unique parameter estimate for the relationship between network density and perceived support adequacy (1 df).
2. All groups were allowed unique estimates for the association between sex composition of support networks and satisfaction with support (3 df).
3. Black males were allowed a unique parameter estimate for the association of density and sex composition (1 df).

The goodness-of-fit indexes for the group-specific model support the removal of these equality constraints. All three indexes indicate that the group-specific model provides an excellent fit to the observed data with little significant covariance left to be explained. Removing these five equality constraints significantly improved overall model fit, $\Delta\chi^2(5, N = 214) = 27.34$, $p < .001$—a significantly better fit between the observed and modeled covariance matrices for all subgroups. The value of the nonnormed fit index fell outside its usual range because of sampling fluctuation as a result of the small sample on which the analyses are based. The comparative fit index, which is more reliable in smaller samples, indicates an excellent fit of these data with the group-specific model.

From the perspective of the fit indexes, the group-specific model in which the effects of three network variables are mediated through perceived social support is consistent with the data. Satisfaction with support was a strong predictor of lower levels of depressive symptoms for all four groups. However, network characteristics predicted satisfaction only for women. This is reflected in individual parameter estimates as well as in larger R^2 for the equation predicting satisfaction.

Relationships Among Network Characteristics. Correlations among the three network variables were positive and significant, with the exception of network density and sex homophily in the black male sample ($r = -.04$). The contrast among the groups on this specific relationship was most striking when black males were compared to white males ($r = .45$).

Network Characteristics as Predictors of Satisfaction With Support. Direct effects of network variables on perceived adequacy of social support were more distinctive than were relationships among network characteristics. None of the parameter estimates of the direct effects of structure on satisfaction for males of either race was significantly different from zero. In contrast, dense networks and networks with opposite sex network members predicted satisfaction with support in the white female sample. The relationship between sex-composition of network and satisfaction was reversed in the black female sample; female-dominated networks were associated with greater satisfaction with support for black women.

Satisfaction With Support and Depressive Symptom Severity. Perceived social support predicted depression for all groups, although the estimate of this relationship varied by almost .10.

DIRECT EFFECTS OF NETWORKS ON DEPRESSIVE SYMPTOMS

To evaluate the assumption that direct effects of support structure on depressive symptoms are unimportant, a final model was tested (parameters not reported). This direct-plus-mediated-effect model freed each group to have unique estimates of the direct effects of network structure variables on depressive symptoms beyond those captured by the mediating effects through perceived adequacy of support. Twelve new parameters were tested for their ability to improve the fit to the observed data.

Out of the twelve new parameters, z-tests indicated that only two were statistically different from zero. For white females, there was a positive relationship between network density and depression that was significant ($b = .27$, $z = 1.98$, $p < .05$), whereas for black females the direct relationship between network size and depression was statistically different from zero ($b = .34$, $z = 2.34$, $p < .05$). For white women, the more dense their social network, the more depressive symptoms they reported; for black women, the larger their social network, the more depressive symptoms they experienced, irrespective of their perceived social support. Although removing the constraints that the direct effects were zero reduced the χ^2 associated with the initial model from 24.60 to 13.51, the loss of 12 degrees of freedom made the fit no longer statistically significant ($p = .009$). Moreover, the $\Delta\chi^2$ with 12 degrees of freedom did not indicate a significant statistical improvement in fit by removing the zero constraints on the direct effects.

Discussion

When taken together, differences and similarities observed in the group-sensitive model recommend it for understanding the relationship between perceived adequacy of support and depression, as well as in the ability of social network characteristics to predict satisfaction with support. Moreover, MSEM made it possible to evaluate the adequacy

of a theoretically derived model as well as to test the generalizability of parameter estimates across major social groups. Substantively, this multigroup structural equation model suggests that different causal mechanisms may link social support networks with perceived adequacy of support. These data also demonstrate a universal linkage between perceived social support and severity of depressive symptoms in this chronically stressed sample of low-income rural elders who have arthritis.

Palinkas et al. (1990), in summarizing their findings, argued that the structure of support networks

> is influenced by age, physical disability, and mortality of network members, and by culturally-determined rules that define the individuals and institutions available for support. However, these rules appear to differ for men and women. (p. 441)

Data presented here support and extend these conclusions by showing that race interacts with sex to define the availability and benefit associated with informal supportive relationships. For all groups, satisfaction with support was associated with lower depressive symptom severity scores. Beyond this, race differences within gender and gender differences within race were as common as the larger effects of gender or race on both access to support and its utility for promoting or protecting psychological well-being. It may be that some groups must rely more heavily on ties that are stressful as well as supportive. If so, the hypothesized benefits of network size, density, and sex composition would be attenuated for these groups relative to other groups.

These findings must be replicated in other samples before they are used to direct policy decisions. The sampling framework for this study was limited and nonrepresentative. In addition, the snowball technique used to generate the sample was not a random process and may oversample unusual subgroups. Generalization of these results beyond sex-race comparisons of low-income rural Southern elders who have arthritis is inappropriate, unless the results are first replicated on more representative samples. In addition, the small samples studied, coupled with the use of post-hoc inductive model fitting techniques, and the large number of parameters estimated in these models are known contributors to increased likelihood of Type I errors. To ensure that these results are not serendipitous, they must be replicated on independent samples.

In conclusion, past research anticipated that the subjects in this sample would be particularly vulnerable to psychological distress if their support systems were inadequate. However, because physical health status, functional limitations, income, and age were controlled through sampling, the usual explanation that race or sex differences are artifacts of disadvantages in health or economic status is untenable here. Observed differences in the amount of available support, as well as in the effectiveness of support for protecting or promoting psychological well-being, emphasize the need for a more sophisticated theoretical and methodological appreciation of the impact of sex and race on social worlds. In particular, we need theories integrating social support and social stratification that recognize structural constraints on well-being that sex and race create, and how these operate throughout the life course. Future theories need to address sex and race independently and interdependently as determinants of psychologically protective social environments, so that formal support programs and policies can be developed to meet the needs of chronically stressed elders.

Multigroup structural equation models serve as powerful tools for advancing theory of the roles of sex and race in the psychological effects of social relationships. Rather than accepting a "one-size-fits-all" theory or relying on cumbersome and often uninterpretable complex interaction models, MSEM provides a direct method for simultaneous testing and evaluating of hypotheses about group effects. MSEM also can be used to estimate group-specific measurement error in latent structure equations, including those based on categorical observed variables. The diversity of MSEM estimation techniques available invites more use of these models for theory building and testing in social and behavioral research.

12 Modeling the Relation of Personality Variables to Symptom Complaints

The Unique Role of Negative Affectivity

JAY G. HULL

JUDITH C. TEDLIE

DANIEL A. LEHN

The last decade has witnessed a proliferation of personality constructs theorized to be relevant to health and health behaviors. These include negative affectivity (Watson & Pennebaker, 1989), self-esteem (Antonucci, Peggs, & Marquez, 1989), social anxiety (Snyder, Smith, Augelli, & Ingram, 1985), sense of challenge (Hull, Petterson, Kumar, & McCollum, 1993), optimism (Scheier & Carver, 1985), cynical hostility (Dembroski & Costa, 1987), hardiness (e.g., Hull, Van Treuren, & Virnelli, 1987; Kobasa, 1979), perceived self-efficacy (Bandura, O'Leary, Taylor, Gauthier, & Gossard, 1987), attributional style (Peterson, Seligman, & Vaillant, 1987), and Type A coronary-prone behavior pattern (Booth-Kewley & Friedman, 1987). When confronted with such a list, the question naturally arises: To what extent are all of

AUTHORS' NOTE: Portions of this chapter were presented at the annual meeting of the American Psychological Society, Washington, DC, June, 1991.

these variables related to health because they are tapping into the same general construct? This question involves two separate issues: (a) is there a general construct that underlies many health relevant personality traits and (b) is it the general trait rather than unique aspects of the individual traits that is related to health outcomes?

General Personality Factors and Their Relation to Health

Recently, considerable progress has been made in specifying a general model of the structure of personality traits. According to this model, personality traits share a five-factor structure (e.g., John, 1990). Typical labels for these factors are neuroticism, extraversion, conscientiousness, agreeableness, and culture. If personality traits have a general structure, then which dimensions of that structure are most strongly related to health behaviors and outcomes? The answer most likely depends on the health outcome of interest. For the present investigation, the domain of principal interest is symptom complaints. Many have suggested that symptom complaints are related to the general factor of neuroticism. This argument has been made both in general reviews of the literature (e.g., Costa & McCrae, 1987; Watson & Pennebaker, 1989) and in research on alternative explanations for the relation of specific personality characteristics to symptom complaints (e.g., as an alternative to hardiness, Funk & Houston, 1987; as an alternative to optimism, Smith, Pope, Rhodewalt, & Poulton, 1989). Others have argued that illnesses are related to variables that are easily recognized as components of neuroticism (e.g., anxiety and depression).

General Factors or Unique Traits?

If one adopts the view that there is a general factor underlying the personality characteristics related to symptom complaints, then one needs to test whether or not a general factor does underlie such traits and whether it is this general factor rather than the unique traits that is responsible for the association of the individual traits with health complaints.

TESTING FOR THE PRESENCE OF A GENERAL FACTOR

Using a confirmatory factor-analytic approach, it is possible to compare different models of personality traits in order to address the question of whether a single factor is adequate to capture the interrelations of a collection of personality characteristics. The models of greatest interest are (a) a first-order one-factor model, (b) a higher-order factor model with one second-order factor, and (c) a group factor model.

According to a *first-order one-factor model*, all of the items from all of the various trait scales are measuring the same thing. More formally, the reason that items from a particular scale are highly correlated with each other (interitem reliability) and the various scales are highly correlated with each other is that all of these measures assess the same construct (e.g., neuroticism).

In contrast to the first-order one-factor model, according to a *higher-order factor model* the individual scales are associated with discriminable dimensions of a higher-order construct. More formally, the reason that items from a particular scale are highly correlated with each other is that they all measure the same, somewhat specific construct. The reason that these specific constructs correlate with each other is that they are all related to a single, general construct. Despite sharing a significant amount of variance with this general construct, the specific constructs retain a significant amount of variance that is unique (unshared with the general construct).

Finally, according to a *group factor model*, individual scales are associated with discriminable dimensions that are linked in complicated, although theoretically sensible, ways. As a consequence of this complexity, considerable variation exists in the degree to which the specific constructs are related. A single, general construct is inadequate to account for this variation.

Each of these models can be subjected to an explicit statistical test using structural equation modeling techniques (e.g., Bollen, 1989b). Such models are typically evaluated in three ways: (a) using overall fit statistics to judge the general adequacy of a particular model, (b) using component fit statistics to judge the adequacy of individual aspects of a particular model, and (c) using overall fit statistics and rules of parsimony to compare alternative models to each other (see Chapters 2 and 5). An acceptable model should show statistically significant components and good overall fit according to a variety of statistics. Direct

comparison should reveal the best model among those hypothesized. Among models that fit equally well, preference is given to the most parsimonious (see Chapter 7). In structural equation modeling, parsimony is judged according to the number of parameters that must be estimated (e.g., Anderson & Gerbing, 1988). In general, first-order factor models are more parsimonious than higher-order factor models, which are in turn more parsimonious than group factor models. If, in fact, a general factor underlies a variety of personality characteristics, then a first-order or higher-order factor model should fit the data as well as a group factor model and should be preferred on the basis of parsimony. Even in cases in which a group factor model fits better than a first-order or higher-order factor model, the latter models might be preferred when a small improvement in fit is gained at the cost of a considerable loss in parsimony.

TESTING FOR THE EFFECTS OF GENERAL FACTORS AND UNIQUE TRAITS

Given evidence that a general factor underlies a variety of health-relevant personality characteristics, it is necessary to test whether the general factor or the unique aspects of individual traits is responsible for an association with health outcomes. Let us take the Watson and Pennebaker (1989) proposition that a variety of personality traits exist that are all related to symptom complaints because they are indicators of neuroticism. This hypothesis is represented in Figure 12.1 as Model 1. In Model 1, each trait (t) is responsible for variation in responses to a number of questionnaire items (i). Variance in the items that is not due to the trait is considered to be a consequence of measurement error (e). At the same time, each trait is associated with the other traits in the model because of their common association with the general latent variable of neuroticism (N). Although each trait is related to N, each also retains some unique variance (u). This unique variance can be thought of as variance in the trait that is unrelated to N but is not simply measurement error. In Model 1, it is the general factor N (rather than the unique aspects of the individual traits) that is responsible for variation in symptom complaints (SC).

Although Model 1 represents our initial theory, let us imagine that we are wrong: Trait t_x is related to symptom complaints because of something unique about it. This can be represented by Model 2 (individual items and errors are not represented in subsequent models for

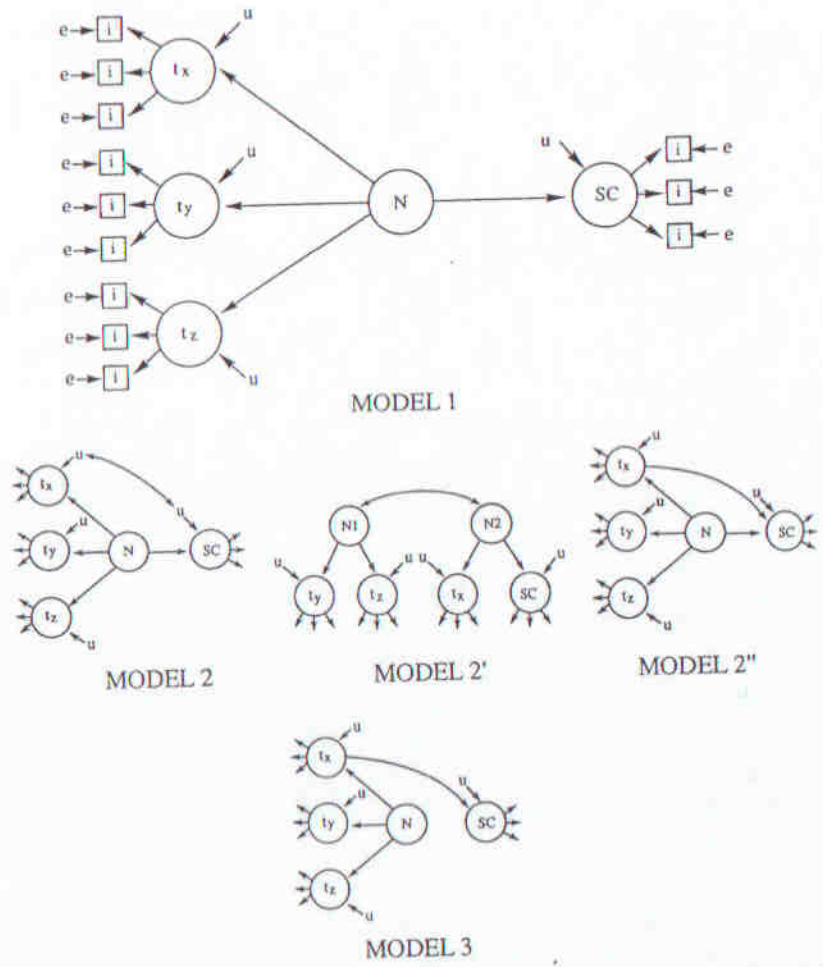

Figure 12.1. Five Models of the Relations Among Personality Traits, General Neuroticism, and Symptom Complaints

the sake of simplicity). In this case, the uniqueness (u) of trait t_x is related to the uniqueness (u) in symptom complaints (i.e., variance in symptom complaints that is not explained by N). Two additional models are statistical (although not conceptual) equivalents to Model 2 and deserve consideration. Model 2′ holds that trait t_x and symptom com-

plaints are both indicators of their own factor (N2) that is correlated with neuroticism (N1). Model 2″ holds that N and trait t_x are independently responsible for variance in symptom complaints.

Finally, let us consider the case in which it is trait t_x and not the general factor N that is responsible for symptom complaints. This possibility is represented in Model 3. Model 3 is a slightly simplified version of Model 2″ in which the direct path between N and symptom complaints has been dropped. According to this model, trait t_x can be thought of as mediating the association between neuroticism and symptom complaints. Thus t_x is either *the* aspect of neuroticism responsible for symptom complaints or it is a separate construct that is affected by neuroticism and in turn affects symptom complaints.

Our own preliminary research suggested that, consistent with the reasoning of Watson and Clark (1984) and Watson and Pennebaker (1989), many traits are related to symptom complaints because of their association with a general factor that we labeled neuroticism. However, one particular trait appeared to play a unique role in predicting symptom complaints. Thus negative affectivity, or a general tendency to report negative emotions, was more highly related to symptom complaints than its association with general neuroticism would suggest. We therefore conducted the three studies reported in this chapter to explore the generality of this effect. In each of these studies, we tested the five models illustrated in Figure 12.1. To foreshadow our findings: In Model 1 the general factor N was always related to symptom complaints despite the fact that the particular personality traits used as indicators of N varied widely across the three studies. Furthermore, Model 2 with negative affectivity playing the role of trait t_x always fit better than Model 1. Although Models 2, 2′, and 2″ provided equivalent fits, Model 2′ showed evidence of misspecification and in Model 2″ the general factor N was never associated with symptom complaints when a direct relation was postulated between negative affectivity and symptom complaints. Given the latter finding, we were always led to Model 3 as a more parsimonious account of the data.

Studies 1-3

Although our data were collected as parts of three separate studies, each study differing from the other in terms of measurement instruments, we report them simultaneously for the sake of brevity and

clarity. Studies 1-3 differed in terms of the particular personality traits that were measured. In each case, traits were selected that other researchers have argued (a) are associated with neuroticism and (b) predict psychological and physical health. In our view, Watson and Clark (1984) have made the strongest argument that a variety of apparently diverse personality scales are in fact measures of the same stable and pervasive trait of neuroticism (e.g., "a number of personality measures . . . , despite dissimilar names, nevertheless intercorrelate so highly that they must be considered measures of the same construct," p. 465). In Study 1, four constructs that Watson and Clark (1984) identified as components of this trait were investigated: negative affectivity, self-esteem, sense of challenge/threat, and social anxiety. In Study 2, social anxiety and self-esteem were replaced as indicators of neuroticism with another variable discussed by Watson and Clark (1984): optimism-pessimism. Study 3 included all five measures of neuroticism included in Studies 1 and 2, plus the six neuroticism subfacets from the NEO Personality Inventory (Costa & McCrae, 1985).

Watson and Pennebaker (1989) argued that the latent variable underlying neuroticism traits is strongly associated with symptom complaints. As evidence, they cite numerous studies that have found an association between symptom complaints and measures they identify as indicators of neuroticism. These arguments were the basis of Model 1 in the present studies. According to this model, all of the individual traits that we have chosen for the present studies are related to symptom complaints only because of their association with a single higher-order factor.

Method

SUBJECTS

Subjects in Study 1 were 177 undergraduates. Subjects in Study 2 were 187 undergraduates. Subjects in Study 3 were 168 undergraduates. All subjects participated in one and only one of the three studies. In each case, subjects participated in return for extra credit in an introductory psychology course. All subjects completed a multiple questionnaire booklet that included the instruments used in these studies. In each study, some subjects failed to complete all of the items from all of the scales and were excluded from further analysis (11 subjects in Study 1, 16 in Study 2, and 22 in Study 3).

INSTRUMENTS

Study 1. Subjects completed the trait version of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) as a measure of dispositional negative affect (Study 1 negative affect subscale $\alpha = .82$), the Rosenberg Self-Esteem scale (RSE; Rosenberg, 1979) as a measure of global self-esteem (Study 1 $\alpha = .88$), the Self-Consciousness Scale (SCS; Fenigstein, Scheier, & Buss, 1975) as a measure of dispositional social anxiety (Study 1 $\alpha = .72$), and the Life Situations Scale (LSS; Hull et al., 1993) as a measure of dispositional challenge (Study 1 $\alpha = .86$). Finally, subjects completed the Cohen-Hoberman Inventory of Physical Symptoms (CHIPS; Cohen & Hoberman, 1983) as a measure of symptom complaints (Study 1 $\alpha = .88$).

Study 2. Subjects completed the Life Orientation Test (LOT; Scheier & Carver, 1985) as a measure of dispositional optimism (Study 2 $\alpha = .82$). As in Study 1, negative affectivity was assessed using the negative emotion traits from the PANAS (Watson et al., 1988; Study 2 $\alpha = .83$), challenge was assessed using the LSS (Hull et al., 1993; Study 2 $\alpha = .84$), and symptoms were assessed using the CHIPS (Cohen & Hoberman, 1983; Study 2 $\alpha = .81$).

Study 3. Subjects completed the NEO-PI (Costa & McCrae, 1985) as a measure of the five major factors of personality. The neuroticism factor is itself divided into six subfacets: self-consciousness, impulsivity, anxiety, depression, hostility, and vulnerability. In the present study, these subfacets were each treated as indicators of neuroticism and were associated with acceptable reliability coefficients (self-consciousness, $\alpha = .74$; impulsivity, $\alpha = .71$; anxiety, $\alpha = .87$; depression, $\alpha = .83$; hostility, $\alpha = .77$; vulnerability, $\alpha = .79$). As in Studies 1 and 2, negative affectivity was assessed using the PANAS (Watson et al., 1988) as a measure of dispositional negative affect, challenge was assessed using the LSS (Hull et al., 1993), and symptoms were assessed using the CHIPS (Cohen & Hoberman, 1983). As in Study 1, social anxiety was assessed using the SCS (Fenigstein et al., 1975) and self-esteem was assessed using the RSE (Rosenberg, 1979). As in Study 2, optimism was assessed using the LOT (Scheier & Carver, 1985). Reliabilities of all scales were similar to those observed in the previous studies.

Results

The first step in analyzing the results involved parcelling each personality trait scale by assigning one third of the items to each of three indicators. Parcelling has multiple advantages over using individual items as indicators (e.g., Rindskopf & Rose, 1988). Of key concern for the present investigation, parcelling allows for the tripartite separation of scale variance into (a) measurement error, (b) variance resulting from the higher-order latent variable of neuroticism, and (c) uniqueness, or variance associated with the individual scale that is not attributed to measurement error or to the higher-order latent variable. Without parcelling, estimates of (b) are attenuated by error, and (a) cannot be distinguished from (c). Finally, specifying three indicators per latent variable is widely recommended as a means of avoiding problems with identification, negative variance estimates, and nonconvergence (e.g., Bollen, 1989b).

PERSONALITY FACTOR MODELS

Before examining the relation of the selected personality traits to symptom complaints, we tested a variety of measurement models of the personality traits themselves. As outlined in the introduction, we specifically compared (a) a one-factor model in which the parcels for all of the individual traits all loaded on a single factor, (b) a higher-order factor model in which parcels for the individual traits loaded on specific trait factors and the latter all loaded on a single higher-order factor, and (c) a group factor model in which parcels for the individual traits loaded on specific trait factors, but the latter correlated in complex ways that were not adequately captured by a single higher-order factor structure. Overall fit was evaluated using χ^2, the Tucker and Lewis (1973) index (TLI), and Bentler's (1990) comparative fit index (CFI). The significance of individual model components was estimated using critical ratios.

One-Factor Model. In Study 1, the one-factor model was associated with 54 degrees of freedom (78 observed variances and covariances, 24 estimated parameters); in Study 2, the one-factor model was associated with 27 degrees of freedom (45 observed variances and covariances, 18 estimated parameters); in Study 3, the one-factor model was associated with 495 degrees of freedom (561 observed variances and covariances,

66 estimated parameters). According to overall fit criteria, the one-factor model resulted in a poor fit to the data in all three studies: Study 1, $\chi^2(54, N = 166) = 341.89$, $p < .001$, TLI = .64, CFI = .71; Study 2, $\chi^2(27, N = 171) = 254.91$, $p < .001$, TLI = .63, CFI = .72; Study 3, $\chi^2(495, N = 146) = 1418.40$, $p < .001$, TLI = .62, CFI = .64 . In each case the model resulted in statistically significant ill fit and accounted for an unacceptably low percentage of the observed covariation.

Second-Order Factor Model. According to overall fit criteria, the second-order factor model resulted in a reasonable fit of the data in Study 1, $\chi^2(50, N = 166) = 99.56$, $p < .01$, TLI = .93, CFI = .95, and Study 2, $\chi^2(24, N = 171) = 33.00$, $p = .10$, TLI = .98, CFI = .99. The data provided a more modest fit of the substantially larger model in Study 3, $\chi^2(484, N = 146) = 835.49$, $p < .001$, TLI = .85, CFI = .87. The factor loadings of the individual parcels on the first-order factors were all sizable and statistically significant. The factor loadings of the first-order latent variables on the second-order latent variable were also all sizable and statistically significant. Furthermore, despite their association with this second-order factor, each of the first-order factors also retained a significant amount of variance that was unique.

Group Factor Model. According to overall fit criteria, the group factor model resulted in a reasonable fit of the data in all three studies: Study 1, $\chi^2(48, N = 166) = 98.56$, $p < .01$, TLI = .93, CFI = .95; Study 2, $\chi^2(24, N = 171) = 33.00$, $p = .10$, TLI = .98, CFI = .99; Study 3, $\chi^2(440, N = 146) = 637.36$, $p < .001$, TLI = .91, CFI = .92. Given the nesting hierarchy, the group factor model can fit no worse than the second-order factor model, although it can fit significantly better than the latter (Rindskopf & Rose, 1988). The individual first-order factor loadings were all statistically significant and their standardized values were all very similar to those that appeared in the second-order factor model. In addition, all of the first-order latent variables were significantly intercorrelated in Studies 1 and 2 and 47 out of 55 of these correlations were significant in Study 3.

Model Comparison. Given that these models are nested versions of each other, they can be directly compared. According to these tests, the one-factor model fits significantly worse than the second-order factor model in all three studies: Study 1, $\Delta\chi^2(4, N = 166) = 242.33$, $p < .001$; Study 2, $\Delta\chi^2(3, N = 171) = 221.91$, $p < .001$; Study 3, $\Delta\chi^2(11, N = 146) = 582.91$, $p < .001$. Given the number of variables used in Study 2, the second-order and group factor models cannot be directly compared in that sample. In Study 1, however, the second-order factor model is not significantly different from the group factor model, difference $\chi^2(2, N = 166) = 1.00$, n.s. Therefore, given the logic of parsimony the second-order factor model constitutes the preferred solution in Study 1.

In Study 3, the group factor model fit significantly better than the second-order factor model, difference $\chi^2(44, N = 146) = 198.13$, $p < .001$. Modification statistics indicated that this was in large part due to the strong relation between particular personality traits assessed by individual scales and particular subfacets of the NEO-PI. In nearly all cases, these relations were consistent with the logic of convergent validity. Thus, by relating all scales and NEO-subfacets to a single general factor, the original model underestimated relations between particular scales and those NEO-subfacets with which they were closely related conceptually (e.g., self-consciousness and social anxiety, optimism-pessimism and depression, challenge and anxiety). At the same time, the majority of these scales and NEO-subfacets were strongly related to the higher-order factor postulated in the second-order factor model (8 of 11 loaded in excess of .75). Because of the a priori status of the second-order factor model, its degree of parsimony and preferred status in Study 1, and its sizable factor loadings, it was retained for the purpose of testing relations among personality traits and symptom complaints.

On the basis of these results, it makes sense to think in terms of discriminable personality traits that share a simple underlying structure while retaining their own unique characteristics. The question then becomes, is this general factor responsible for variation in symptom complaints that is observed to be associated with these personality traits or is this variation due to unique characteristics of particular traits?

PREDICTING SYMPTOM COMPLAINTS

The relation of these personality characteristics to symptom complaints was modeled within the framework provided by Model 1 illustrated in Figure 12.1. The measure of symptom complaints was parcelled by assigning one third of the items to each of three indicators. An initial model was then specified in which the latent variable of symptom complaints was only related to the second-order factor of neuroticism. In Study 1, Model 1 was associated with 85 degrees of freedom (120

observed variances and covariances, 35 estimated parameters); in Study 2, Model 1 was associated with 50 degrees of freedom (78 observed variances and covariances, 28 estimated parameters); in Study 3, Model 1 was associated with 582 degrees of freedom (666 observed variances and covariances, 84 estimated parameters).

According to overall fit criteria, Model 1 resulted in a reasonable fit to the data in Study 1, $\chi^2(85, N = 166) = 149.88$, $p < .001$, TLI = .94, CFI = .95, and Study 2, $\chi^2(50, N = 171) = 75.03$, $p = .01$, TLI = .97, CFI = .97. Study 3 involved a considerably larger model and resulted in a slightly poorer fit for Model 1, $\chi^2(582, N = 146) = 967.12$, $p < .001$, TLI = .85, CFI = .86.

As expected, in all three studies, component fit criteria indicated that the second-order variable of neuroticism was significantly related to symptom complaints, standardized path Study 1 = .35, $p < .001$; Study 2 = .39, $p < .001$; Study 3 = .43, $p < .001$. In addition, in each study, modification statistics suggested that this Model 1 could be improved by allowing the residual variance in the latent variable of negative affectivity to correlate with residual variance in symptom complaints. Multivariate modification statistics indicated that in all three studies this was the only personality trait residual related to the symptom complaint residual. As noted earlier, this indicates that the second-order factor model is inadequate to account fully for the variance shared by negative affectivity and symptom complaints. At the same time, it is important to note that the second-order factor model was adequate to account fully for the variance shared by symptom complaints and all of the other personality traits in all three studies.

Model 2 was identical to Model 1 except that it allowed residual variance in negative affectivity to correlate with residual variance in symptom complaints (see Figure 12.1). In each study, the result was a significant improvement in fit, Study 1, $\Delta\chi^2(1, N = 166) = 8.03$, $p < .01$; Study 2, $\Delta\chi^2(1, N = 171) = 7.46$, $p < .01$; Study 3, $\Delta\chi^2(1, N = 146) = 4.33$, $p < .05$. The interpretation of the altered model is as follows: Although it is useful to think of many traits as redundant for the sake of predicting symptom complaints (and these traits vary widely across Studies 1-3), this is not true of negative affectivity. Although negative affectivity is related to the general variable of neuroticism, it is more strongly related to symptom complaints than would be expected given its association with neuroticism.

We considered two additional ways to model the unique relation of negative affectivity and symptom complaints within the framework of

the higher-order factor model of neuroticism. These two models are identical in overall fit to Model 2 and are simply alternative ways of conceiving the observed relations (see Breckler, 1990; Chapter 2, this volume). Model 2′ allowed the personality traits other than negative affectivity to load on one second-order factor (neuroticism 1) and negative affectivity and symptom complaints to load on their own separate second-order factor (neuroticism 2). The two second-order factors were then allowed to correlate. Although this model resulted in the same χ^2 as Model 2, it also yielded a negative residual variance estimate for negative affectivity in both Study 1 and Study 2. Such results may indicate that the model is misspecified (e.g., Bollen, 1989b).

Model 2″ allowed all personality traits to load on the second-order factor of neuroticism; however, both neuroticism and negative affectivity were used as independent predictors of symptom complaints. Again, this model yields a χ^2 identical to that observed for Models 2 and 2′. However, when modeled in this way, the independent association of neuroticism and symptom complaints is nonsignificant in all three studies (standardized path Study 1 = −.20, n.s.; Study 2 = −.29, n.s.; Study 3 = .15, n.s.), whereas the path associating negative affectivity and symptom complaints is significant in all three studies (standardized path Study 1 = .58, $p < .01$; Study 2 = .64, $p < .01$; Study 3 = .35, $p < .05$). Dropping the nonsignificant neuroticism-symptom complaints path results in Model 3. In this model, negative affectivity was significantly related to symptom complaints in all three studies (standardized path Study 1 = .40, $p < .001$; Study 2 = .40, $p < .001$; Study 3 = .48, $p < .001$). Because Model 3 is a nested version of Model 2″, it can be directly compared to it. The resulting χ^2 is not significant in any of the three studies, Study 1, $\Delta\chi^2(1, N = 166) = .94$, n.s.; Study 2, $\Delta\chi^2(1, N = 171) = 1.44$, n.s.; Study 3, $\Delta\chi^2(1, N = 146) = .84$, n.s. Because Model 3 estimates one less parameter, it may be preferred on the basis of parsimony. Model 3 for Study 3 is shown in Figure 12.2.

Discussion

On the basis of these studies, it would appear that (a) a variety of personality traits that have been argued to relate to physical and psychological health might usefully be considered to share a common association with an underlying general factor of neuroticism; (b) despite their association with neuroticism, each of these traits retains a

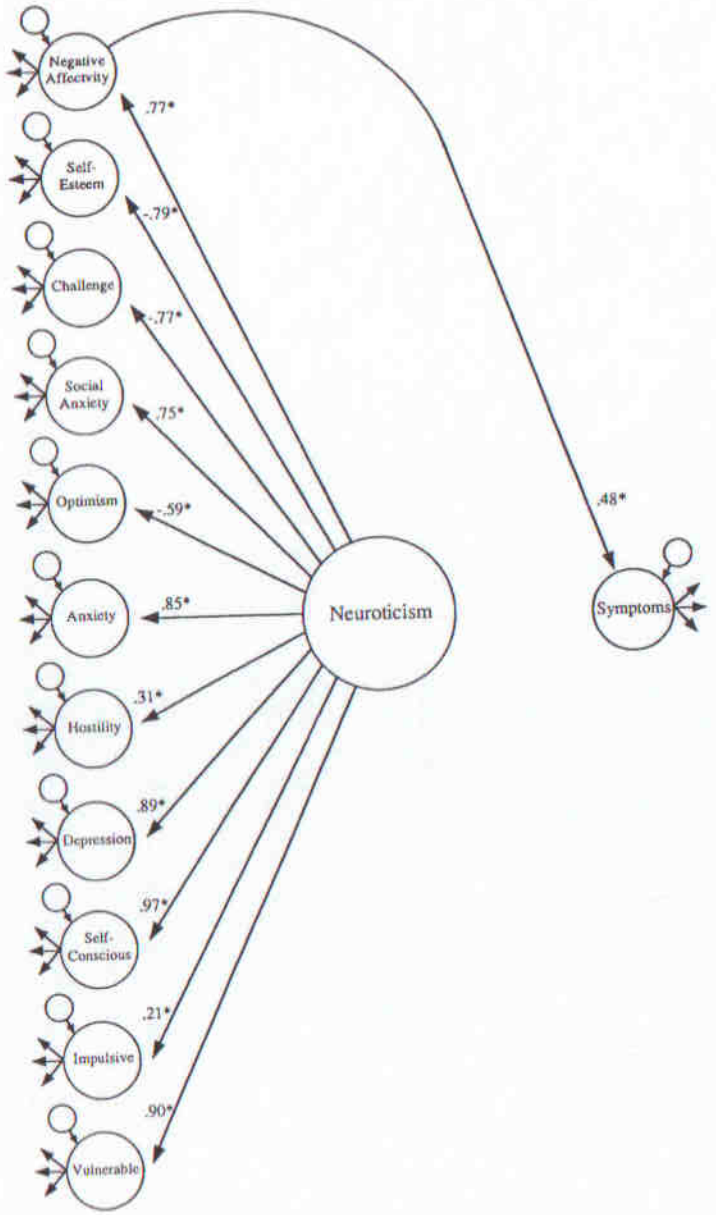

Figure 12.2. Study 3, Model 3: Negative Affectivity as a Mediator Between General Neuroticism and Symptom Complaints

substantial amount of unique variance; and (c) in most cases these personality traits are related to symptom complaints because of their association with general neuroticism rather than because of their unique characteristics; however, (d) negative affectivity is more strongly related to symptom complaints than can be explained by its association with the general factor of neuroticism.

Given that negative affectivity would appear to be ill-conceived as simply an indicator of general neuroticism, we considered a variety of ways to model its association with symptom complaints and the remaining personality characteristics. Model 3 was offered as providing the most parsimonious account. This model has a rather straightforward interpretation: A variety of different personality traits are to some extent overlapping constructs that share a common source of variance with a general factor of neuroticism. These variables are related to each other and to negative affectivity because of their shared association with neuroticism rather than because of anything unique about them as personality traits. Negative affectivity in turn serves to mediate the influence of general neuroticism on symptom complaints.

THE UNIQUE QUALITIES OF NEGATIVE AFFECTIVITY

The results of these studies naturally lead one to ask: What is unique about negative affectivity? Obviously, the answer to this question goes beyond the data at hand. Nevertheless, in comparing personality trait measures used in the present studies, it became apparent to us that they vary in their assessment of subjects' dispositional tendencies to think, feel, and/or act in particular ways. Specifically, some scales are dominated by questions regarding thoughts (e.g., the Rosenberg Self-Esteem Scale, the Life Orientation Test of optimism-pessimism) and others are dominated by questions regarding acts (e.g., the NEO impulsivity subfacet), but most contain a mix of questions regarding thoughts and feelings (e.g., the NEO depression subfacet, the NEO anxiety subfacet, the Life Situation Scale of challenge). In contrast to the other scales, the negative affectivity subscale of the PANAS exclusively asks about subjects' feelings or emotions. Thus one way to view the present results is that general neuroticism is composed of a variety of thoughts, feelings, and behaviors; however, the feeling component (best assessed by the PANAS measure of negative affectivity) is responsible for symptom complaints. From this perspective, thoughts and behaviors give rise to symptom complaints only to the extent that they are associated with negative emotionality.

Why should negative emotions be especially likely to yield symptom complaints? Salovey (1991) has offered several distinct reasons for such an effect. Negative emotions may (a) focus attention inward, leading to increased awareness of all internal states including physical symptoms (see also Watson & Pennebaker, 1989), (b) bias memory and hence reporting of physical complaints, (c) lead to behaviors that damage health (e.g., smoking, alcohol consumption) and hence increased illness, (d) damage the immune system leading to increased susceptibility to illness, or (e) motivate a desire for attention from others and hence increased likelihood to express symptom complaints to elicit sympathy. To this list we would add that physiological changes associated with chronic negative emotions may be confused with symptoms of physical illness.

Regardless of the specific interpretation of the unique association of negative affectivity and symptom complaints, it is apparent that negative affectivity does not operate like other indicators of neuroticism. As a consequence, we recommend that measures be developed that clearly separate affective from nonaffective components of personality so that the effects of these components can be independently observed. Development of such measures is particularly important to the extent that personality variables have effects that are not affectively mediated and trait affectivity has sources that are not personality based (e.g., life circumstances). In addition, separation of affective from nonaffective aspects of personality will yield models whose affective components can be subjected to more rigorous tests of causality through independent manipulation. For example, in research relevant to the current studies, Salovey and Birnbaum (1989) demonstrated that manipulations of negative affect do in fact increase symptom complaints.

METHOD AND KNOWLEDGE: THE UTILITY OF STRUCTURAL EQUATION MODELING

General and Specific Effects. Elsewhere we have detailed the advantages of structural equation modeling approaches to evaluating multidimensional personality constructs (Hull, Lehn, & Tedlie, 1991; Hull & Mendolia, 1991). The principal advantage in the present research has been to identify the existence and predictive utility of the unique variance in different health relevant personality constructs. In all of the models, this was accomplished by dividing the variance in the observed measures into (a) variance resulting from the general, second-order

construct, (b) variance resulting from the unique aspects of the specific construct, and (c) variance resulting from measurement error. Traditional approaches obscure these sources and as a consequence make it difficult to assess the unique versus redundant aspects of personality measures.

In this context, it is worth noting that although the unique aspects of the personality traits other than negative affectivity were not useful in predicting symptom complaints, the estimated size of these variances is large and statistically significant. It is quite reasonable to expect that the unique aspects of these variables will prove useful in predicting variables other than symptom complaints.

Modeling Strategy. In addition to adopting a measurement approach that allows more detailed examination of the source of observed effects, we also used data-analytic strategies that we recommend to anyone using structural equation modeling. First, we adopted a measurement model (the second-order factor model) that tried to balance the twin criteria of quality of fit and parsimony. Second, we examined a variety of a priori structural models (Models 1-3) including some that were statistically identical in fit (Models 2, 2′, and 2″). Third, we attempted to document the robust character of the observed effects by conducting replications in multiple samples while varying the specific indicators of the latent variable of interest.

LIMITATIONS OF THE PRESENT RESEARCH

Although our approach has definite advantages over alternative approaches, the methods used in the present studies also have distinct limitations. These limitations include (a) a select range of neuroticism indicators, (b) the potential for misunderstanding the status of models as evidence, and (c) reliance on symptom complaints as a measure of health behavior.

Neuroticism Indicators. As in any research, we have been forced to rely on a limited number of measured variables to make a more general case. In doing so, we have tried to sample broadly from the variety of measures available. Most of the measures included in our models have been identified as key components of neuroticism in general reviews of the literature. For example, Watson and Clark (1984) explicitly identified self-esteem, negative emotions, pessimism, and a tendency to

perceive the environment as threatening (i.e., low challenge) as aspects of general negative affectivity. Similarly, Costa and McCrae (1987) identified self-esteem, negative affectivity, social anxiety, and a perceived inability to cope (i.e., low challenge) as aspects of neuroticism. In addition to including a wide variety of measures, we have also varied these measures across studies in an attempt to demonstrate that our results are not dependent on a particular combination of items. As a consequence, we are reasonably confident as to the generalizability of our results. Nevertheless, the variety of health relevant personality constructs is large enough that not all constructs may be subsumed under this general model. In particular, constructs that are unrelated to neuroticism need to be identified and validated.

Interpreting the Results of Structural Equation Modeling. Despite its distinct advantages, structural equation modeling does not provide evidence of causality, and it does not "prove" the superiority of one model over all possible alternative models (e.g., Breckler, 1990; Cliff, 1983; Chapters 1 and 2, this volume). With respect to the latter issue, although we have shown that certain models are superior to specific alternatives, we have not eliminated the universe of alternative models.

Measures of Health and Health Behaviors. We have no illusions that by predicting symptom complaints we have demonstrated a link between personality and disease. Many researchers have provided extensive evidence calling into question the use of symptom reports as valid measures of physical illness (e.g., Costa & McCrae, 1987; Watson & Pennebaker, 1989). Most of these researchers see symptom complaints as health relevant behaviors that are to some extent a function of personality neuroticism. It is from this perspective that we view symptom complaints as providing an appropriate context for testing the unique predictive utility of the various subcomponents of neuroticism.

Conclusion

Based on the present research, it is our view that some researchers have moved too quickly toward the view that a variety of health relevant personality constructs are simply surrogates for neuroticism. With respect to the alternative models described in the introduction, the one-factor model is clearly wrong. Furthermore, the single higher-order

factor model, although adequate in some respects, clearly fails to account fully for the association of negative affectivity and symptom complaints. Finally, the group factor model appears to be unnecessarily complex. The model that was ultimately derived and replicated combines a higher-order factor model with a group factor model. According to this approach, some health relevant personality constructs have their influence on symptom complaints via the general factor of neuroticism, whereas others (i.e., negative affectivity) actually serve to mediate the influence of general neuroticism on symptom complaints. Future research should be directed toward determining the aspect of negative affectivity that distinguishes it from other traits related to neuroticism in such a way that it is a uniquely strong predictor of symptom complaints.

13 Predictors of Change in Antisocial Behavior During Elementary School for Boys

MIKE STOOLMILLER

TERRY E. DUNCAN

GERALD R. PATTERSON

Lykken (1993) recently came to some rather startling and discomforting conclusions about the level of antisocial and aggressive behavior in the United States:

> The United States has the highest per capita incidence of interpersonal violence of any nation not actively engaged in civil war. In a comparison of homicide rates (per 100,000 inhabitants) of 20 industrialized countries, the United States not only ranks first but has a rate more than four times greater than the nearest competitor. (p. 17)

The above quotation underscores the importance of understanding the emergence, escalation, and eventual maintenance of serious antisocial behavior.

This chapter utilizes structural equation modeling (SEM) techniques to test a social learning perspective on the development of antisocial

AUTHORS' NOTE: Preparation of this chapter was supported in part by grant MH 46690 from the Prevention Research Branch, NIH; grant MH 37940 from the Center for Studies of Antisocial and Violent Behavior, NIH; grant MH 38318 from the Mood, Anxiety, and Personality Disorders Research Branch, NIH; and grants DA 07031 and DA 03706 from the National Institute on Drug Abuse.

behavior in grade-school boys that has emerged from years of research by Patterson and his colleagues on coercive family interaction (see Patterson, Reid, & Dishion, 1992, for details). Coercion theory emphasizes simple, well-founded principles of positive and negative reinforcement from learning theory to explain the emergence, escalation, and eventual maintenance of serious antisocial behavior in children. Specifically, coercion theory holds that over many thousands of trials during family interaction, the problem child learns via negative reinforcement to shut off irritable, aversive intrusions from parents and siblings with his or her own aversive counterattacks. With parents, these attack-counterattack sequences typically occur during discipline encounters and undermine the parents' abilities to socialize and supervise the child properly. Although it does not emphasize biological or temperamental variables, coercion theory is still consistent with the view of Lykken (1993) that some children may be more susceptible to the detrimental effects of coercive family interaction than others. As the result of an abrasive, aggressive interpersonal style and low parental supervision, the child is rejected by normal peers, fails to learn at school, and eventually winds up in a delinquent peer group as an adolescent (Dishion, Patterson, Stoolmiller, & Skinner, 1991).

It is increasingly clear that antisocial behavioral patterns begin early and are highly stable across development (Loeber, 1982; Loeber & Dishion, 1983). Robins (1978) concluded that nearly all adults with antisocial personality disorder had demonstrated troublesome and oppositional behavior beginning between 8 and 10 years of age. More recent studies strongly suggest that the pattern begins before first grade (Campbell, Ewing, Breaux, & Szumowski, 1986; Loeber & Dishion, 1983). West and Farrington (1977) concluded that about 7% of the adolescent population who were chronically antisocial accounted for between 50% and 70% of all the reported juvenile criminal acts. Thus, on the basis of the continuity and intensity of the problem, identifying potentially malleable predictors of childhood antisocial behavior such as parenting practices could have important implications for efforts to prevent or reduce such behavior.

The focus of this chapter is on expanding the test of key coercion theory variables such as parental discipline skill and monitoring practices. This is accomplished by including other potentially competing explanatory variables in a model of changes in antisocial behavior in the school setting from Grade 4 to Grade 5.

Focusing on changes in antisocial behavior in the school setting as an outcome has several advantages. First, it minimizes any potential biases or distortions that might arise when all information comes from one informant in one setting (e.g., the mother reporting about child behavior at home). Second, it demonstrates that aggressive and antisocial behavior learned at home will generalize or spill over to school settings. Third, it focuses attention on antisocial behavior that would tend to undermine success in school, which in turn carries negative long-term implications for successful adult adjustment.

Emphasizing dynamic change in antisocial behavior from Grade 4 to Grade 5 as opposed to just status at Grade 5 also strengthens the test of coercion theory. It avoids the implicit, untested, and unreasonable assumption that predictors of Grade 5 antisocial status are equivalent to predictors of change. This assumption will hold only when antisocial behavior essentially emerges de novo in the school setting at Grade 5 or change is completely determined by Grade 4 status. Neither of these conditions seems plausible. A focus on models of change enables an examination of how key coercion theory variables exert an influence on an ongoing process.

Studying competing predictors also strengthens the test of coercion theory. Meehl (1990) has argued persuasively that, in general, rejecting the statistical null hypothesis of no effect is a very weak test of a substantive theory in psychology. Thus a large number of other plausible explanatory variables will be examined for their independent contribution to the prediction of changes in antisocial behavior. Grade 4 characteristics of the child such as self-esteem, depressed moods, and peer relations; parent characteristics such as depression, antisocial tendency, irritability, and marital adjustment; and family characteristics such as parental socioeconomic status, family income, and family problem solving skill will all be included. In addition, the key coercion theory variables of parent discipline skill, parent monitoring practices, and child coerciveness will be included.

The hypothesis to be examined for this study is that key coercion theory variables will exert significant, direct effects on two measures of change in antisocial behavior at school, even when controlling for the effects of the competing variables. The statistical model that will be utilized will be a measured variable path model that is essentially a simultaneous multiple regression model with two correlated outcome measures.

Method

SAMPLE

The sample used for these analyses comes from two successive Grade 4 cohorts of the Oregon Youth Study (OYS) being conducted by Patterson and his associates. Schools in the metropolitan Eugene-Springfield area of Oregon were ranked from highest to lowest by police statistics for rates of delinquency. For Cohort I, six schools were randomly selected from the top 10 and all families with a boy in Grade 4 were invited to participate. For Cohort II, recruited one year later, seven schools were randomly selected from the top 10. This scheme generated a sample of 206 families (102 for Cohort I and 104 for Cohort II), for an overall participation rate of 74.4%. The sample is primarily European-American (86%), and a majority of the boys come from two-parent families (70%). For these analyses, 10 single-father families were excluded because of the small sample size, which brought the total sample size to 196. At least 50% of the families in both cohorts were of either low socioeconomic status or working class. The families were paid up to $300 for participating in the annual assessments.

INSTRUMENT DESCRIPTION

The OYS annual assessment battery includes structured interviews and telephone interviews with the study boy and his parents; questionnaires administered to the boy, his parents, and the boys' teachers; three home observation sessions lasting one hour each; and a structured family problem solving task completed in the lab. Data for these analyses were taken from the initial assessment at Grade 4 and again one year later at Grade 5.

To complete the telephone interview, each family is called six times over approximately three weeks. Items for the telephone interview consist of the occurrence of behaviors of interest either in the last 24 hours or in the last three days. In all cases, the average score from the six calls is used as the basic item in scales.

Each family was observed three times over approximately three weeks for one hour each time in order to complete the home observations. In addition, home observers filled out observer impressions inventories immediately after each observation session. Observations were conducted in the late afternoon near the dinner hour. Behavior was

recorded during home observations using the Family Process Coding (FPC) system, which was described in detail by Dishion et al. (1983).

In general, scales were created separately for each unique combination of respondent and assessment method. A priori item pools were created on the basis of strong face validity. Items were standardized prior to item analyses to ensure that arbitrary scaling differences across respondents and instruments did not influence the total variance of the construct score. Item analysis was conducted on the a priori item pools; items with nonsignificant, corrected item-to-total correlations were dropped from the scales. Scales had to produce a coefficient of internal consistency above .60 or they were dropped. Scales were computed if at least 60% of the items were present; otherwise, the score was coded as missing. All scales for a given respondent were averaged to form a respondent-specific indicator of the construct. Finally, construct scores were computed by averaging over all the respondents, given that at least 60% of the respondent-specific indicators were present. Computing constructs by aggregating across methods and respondents minimizes any bias unique to a method or respondent that would tend to limit the generalizability of the findings.

CHILD CONSTRUCTS

School antisocial behavior at Grade 4 was obtained from teacher ratings using the CBC-L (Achenbach & Edelbrock, 1986) and peer nominations for aggressive and troublesome behavior. Peer nominations were standardized within classrooms. At Grade 5, two school antisocial construct scores were derived. The first was based solely on teacher ratings on the CBC-L and was available for all 196 subjects. The second was a much richer measure that included direct observation of behavior in the classroom (amount of time on task) and on the playground (negative social interaction), discipline contacts, attendance, and a teacher rating of positive school adjustment from the Walker-McConnell scale of social competence and school adjustment (Walker & McConnell, 1988). Unfortunately, the additional school antisocial measures were obtained for only a subsample ($n = 79$) of the OYS. Therefore, SEM techniques for missing data (described in greater detail in the modeling section) were utilized in order to include both antisocial outcome measures in a single model.

The *playground negative social interaction* measure was derived from the Target/Peer Interaction Code (TPIC; O'Neil, Ramsey, Shinn,

Todis, & Spira, 1985), which recorded the free play and social behavior of the target subjects and their peers in playground settings. The playground measure is a combination of the total negative behaviors by the target subject and the total negative behaviors by peers directed at the target subject. *Academic engaged time* was assessed within each subject's classroom using a definition and duration recording procedure developed by Walker, Severson, Haring, and Williams (1986). A student was defined as academically engaged if he was attending to the assigned material and the academic tasks involved, was making appropriate motor responses (e.g., writing or computing), or was appropriately asking for assistance in an acceptable manner. Direct observation of engaged time was conducted during two 15-minute sessions in which reading or math was being taught and the students were expected to do independent seat work. *Attendance and discipline contacts* with the school principal were obtained at the end of the school year from official school records. Change in school antisocial behavior for these analyses refers to residual change; that is, the part of Grade 5 school antisocial behavior that is not accounted for by Grade 4 school antisocial behavior.

Tantrums and disobedience was derived from parents' ratings on the CBC-L (Achenbach & Edelbrock, 1983), the parent interview, and parent telephone interview. Scores were first derived for tantrums and disobedience separately and then combined so that each would contribute more or less equally. Separate ratings of *depressed mood* and *good peer relations* were obtained from parents' and teachers' ratings on the CBC-L. *Self-esteem* was based on self-report using Rosenberg's (1965) Self-Esteem Scale.

PARENT CONSTRUCTS

Depression was derived from the CES-D (Radloff, 1977). *Irritability* was based on observers' impressions of observed irritability during the home observations. *Antisocial* was computed from arrest records obtained from the Oregon Department of Motor Vehicles and from FBI records. *Marital adjustment* was based on the proportion of behavior by one partner directed toward their spouse that was coded as aversive using a priori aversive categories of the FPC. *Good discipline* includes both observed rates of behavior and observer impressions items pertaining to discipline practices and lack of discipline control of the boy. The observed rates included nattering, the proportion of the parents' behav-

ior directed at the child that was coded as low intensity aversive (i.e., not physically aggressive), and abusing (i.e., physically aggressive) using a priori aversive categories of the FPC. *Inept monitoring* was based on separate interviewers' impressions for the mother, father, and child concerning the adequacy of parental monitoring and supervision of the study boy. *Problem solving* was based on a structured problem-solving task that was conducted in the lab and includes the total number of positive solutions suggested by family members, a binomial *z*-score reflecting the tendency of the parents to encourage the child's participation, and observers' impressions concerning the quality of the problem-solving process and proposed solutions. *SES* was scored using the method of Hollingshead (1975) and includes educational attainment and occupational prestige. *Poverty* is computed by dividing the total family income by the number of family members, standardizing over the sample, and reversing the scaling. *Family type* is a nominal code indicating at Grade 4 whether the family is a two-parent family ($n = 54$ with complete data, $n = 86$ with missing Grade 5 data) or a single-mother family ($n = 25$ with complete data, $n = 31$ with missing Grade 5 data).

MODELING DESCRIPTION

Some additional description of modeling techniques is warranted because of the additional complexity involved in handling missing data on the Grade 5 school antisocial measure. A subsample of 79 boys was selected from the larger sample of 196 for more intensive study in the school setting. Selection was based on a global measure of antisocial behavior obtained at Grade 4 that included parent, peer, and teacher ratings and direct observation in the home and lab. The 40 most antisocial boys on the global index were included along with a random sample of 39 from the remaining 166 boys. Clearly, the subsample of 79 cannot be considered a completely random sample from the larger sample of 196. However, the use of SEM techniques for missing data can proceed under the much milder assumption that the missing data at Grade 5 are missing at random (MAR) once the sampling mechanism is accounted for (Little & Rubin, 1990; Muthén, Kaplan, & Hollis, 1987). In other words, the school antisocial data that are missing at Grade 5 are not systematically related to the (unobserved) level of school antisocial at Grade 5 once the Grade 4 selection criteria are considered. In the terminology of Little and Rubin (1990), the missing data mechanism is said to be ignorable and the techniques suggested by Allison and Hauser (1991), Bentler (1989), and Muthén et al. (1987) can be used. Interested readers should refer to the above references for more details and sample programs.

Briefly, the strategy for handling missing data consists of expanding the usual SEM model to include means and intercepts, and partitioning the sample into subgroups with distinct patterns of missing data. Equality constraints across the missing data groups are used in a multisample analysis to obtain unbiased, consistent estimates. It should be emphasized that these equality constraints across the missing data subsamples are not substantively interesting; they function solely to ensure correct estimation of model parameters. For these data, preliminary regression analyses indicated that there were also significant differences in the models across family-type groups (two-parent vs. single-mother). Thus the sample was split into four separate groups defined by family type (single-mother vs. two-parent) and missing data (complete vs. missing data on additional Grade 5 school antisocial measures). Equality constraints across the missing data groups within family-type groups were employed to obtain correct parameter estimates.

Results

Regression imputation was used separately in the two-parent and single-mother subsamples to replace a small number of missing values that were scattered over the data (eight subjects missing one value each). Missing values on the predictor variables were estimated from regressions using the other predictors with complete data. One missing value on the Grade 5 teacher CBC-L ratings of school antisocial behavior was estimated using the Grade 4 school antisocial behavior measure.

The teacher ratings of school antisocial behavior showed a significant mean level increase from Grade 4 to Grade 5. Boys in stepparent families increased the most relative to boys in intact or single-mother families, although the difference was only marginally significant at the .07 level.

Separate, preliminary regression analyses for each Grade 5 outcome were conducted using the entire sample ($N = 196$) for the Grade 5 teacher ratings of school antisocial behavior and the subsample ($n = 79$) for the general school antisocial behavior measure. The purpose of

these analyses was to reduce the number of potential predictors and to examine the data for outliers and highly influential observations. The following variables were dropped from the models because they had no impact on either Grade 5 outcome measure when controlling for other predictors: depressed mood (both parent and teacher ratings), self-esteem, good peer relations (parent ratings), irritability, marital adjustment, SES, and problem solving. Variables retained for further consideration either had significant effects in the preliminary analyses or were considered key theoretical variables.

Because the model includes an interaction term, Figure 13.1 shows estimated, unstandardized regression weights for the two outcome variables and means and variances for the predictor variables for single-mother versus two-parent families with no constraints across family-type groups. The overall model χ^2 is 194.63 with 130 degrees of freedom (df) and a p-value less than .001. The large χ^2 value and df result from the constraints across missing data groups. If there were no missing data, the model shown in Figure 13.1 would be completely saturated, using up all available df and fitting the data perfectly. The missing data constraints represent a test of the hypothesis that the data are missing completely at random. As mentioned earlier, this is not substantively interesting because the sampling mechanism is known and is definitely not a simple random sampling scheme. The obtained χ^2 value is useful, however, as a baseline for testing more substantively interesting hypotheses about the equality of regression weights across family-type groups. The df for each family-type group can be obtained by noting that the group with missing data will contribute 65 df (45 covariances, 10 variances, and 10 means), and the group with no missing data will contribute 77 df (55 covariances, 11 variances, and 11 means) for a total of 142 df. The model being fit has 77 estimated parameters: 54 covariances, variances, and means for the 9 predictors; 1 residual variance and 1 intercept for the outcome with no missing data; 18 regression weights linking the predictors to the outcomes—all of which are constrained to be equal across missing data groups—and, finally, 1 residual variance, 1 intercept, and 1 residual covariance for the outcome that has missing data. Subtracting the 77 df for the 77 estimated parameters from the total available 142 df leaves 65 df per family-type group. For the model with no constraints across the two family-type groups, there are 284 total available df (142 df per family-type group × two family-type groups) and 154 estimated parameters, leaving the model df equal to 130.

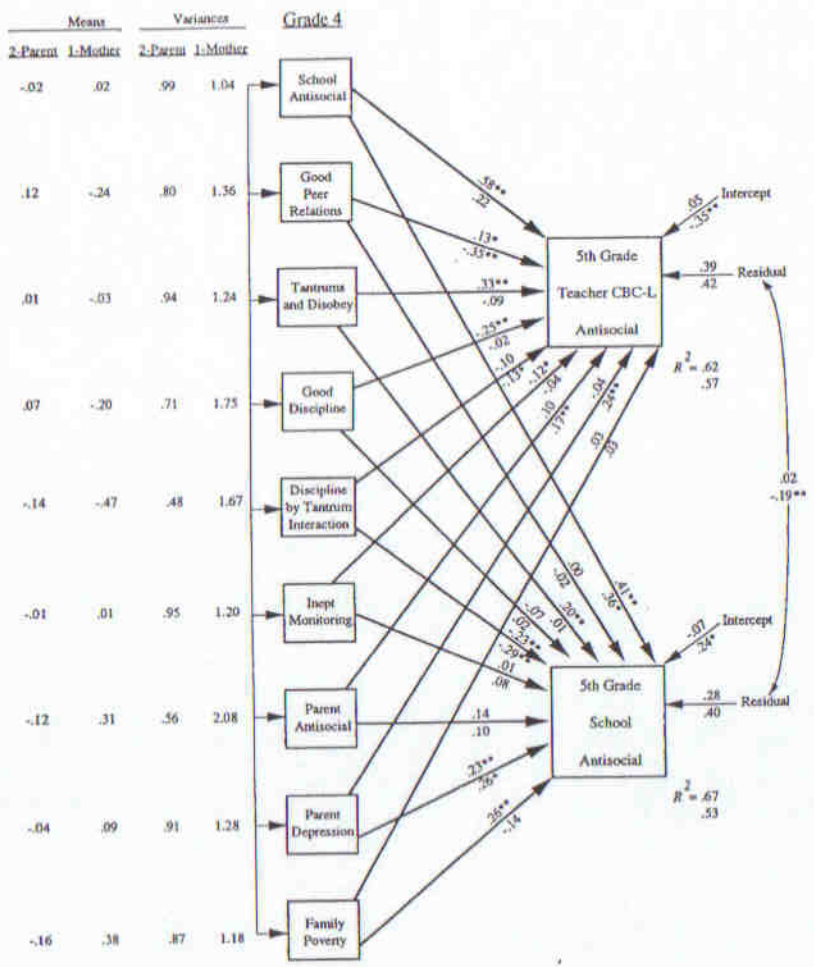

Figure 13.1. Change in School Antisocial Behavior, No Constraints Across Family Type

NOTE: Estimates are unstandardized. Two-parent family values shown on top, single-mother values on bottom.
*p < .05, **p < .01.

For the model with complete constraints across family-type groups, the total available df is still 284 but now only 77 parameters are estimated, leaving the model df equal to 207 with a χ^2 of 421.79 and a

p-value less than .001. The difference in χ^2s between the full constraints model and the no constraints model can be used as a statistical test of the hypothesis that the models are the same in two-parent and single-mother families. This $\Delta\chi^2$ is equal to 227.16 with 77 df and a p-value less than .001. The test indicates that some aspect(s) of the model differ significantly across the two family-type groups. The comparative fit index (CFI) for comparing the completely constrained versus the unconstrained models is a substantial .699.

In order to explore differences in the model across family types, modification indices were used to identify 16 constraints across family-type groups that seemed untenable. Once these 16 constraints were relaxed, the model generated a χ^2 of 263.80 with 191 df for a p-value less than .001. The test of the significance of the 16 relaxed constraints was obtained by subtracting this χ^2 from the full constraints model χ^2 to obtain a $\Delta\chi^2$ of 157.99 with 16 df and a p-value less than .001. The CFI corresponding to this comparison is .661. Clearly, relaxing the 16 constraints results in a large and significant improvement in fit of the model. A test of the tenability of the remaining constraints is obtained by subtracting the no constraints model χ^2 from the partially constrained model χ^2 to obtain a $\Delta\chi^2$ of 69.17 with 61 df and p-value greater than .05. The CFI for this comparison is only .112. Thus the rest of the constraints can be considered reasonable. Most of the model parameters (61 out of 77) can be considered the same across the two family-type groups. Note that results concerning differences in specific parameter values across family types should be considered exploratory, and significance tests should not be taken too seriously.

Unstandardized regression weights for both outcomes and means and variances for the predictors for the partially constrained model are shown in Figure 13.2. Different values in Figure 13.2 across family groups indicate significant differences in parameter estimates across family groups. Single-mother families were significantly higher on both poverty and depression. Single-mother families also had significantly larger variances on good discipline, the interaction term of tantrums and disobedience with good discipline, antisocial, and poverty.

The effects of the predictors were much more variable across family groups for teacher ratings than for general school antisocial behavior. Seven of the eleven estimated parameters for the teacher ratings were different across family groups as compared to only one of the eleven for general school antisocial behavior. In fact, for the teacher ratings of school antisocial behavior, teacher ratings of good peer relations had

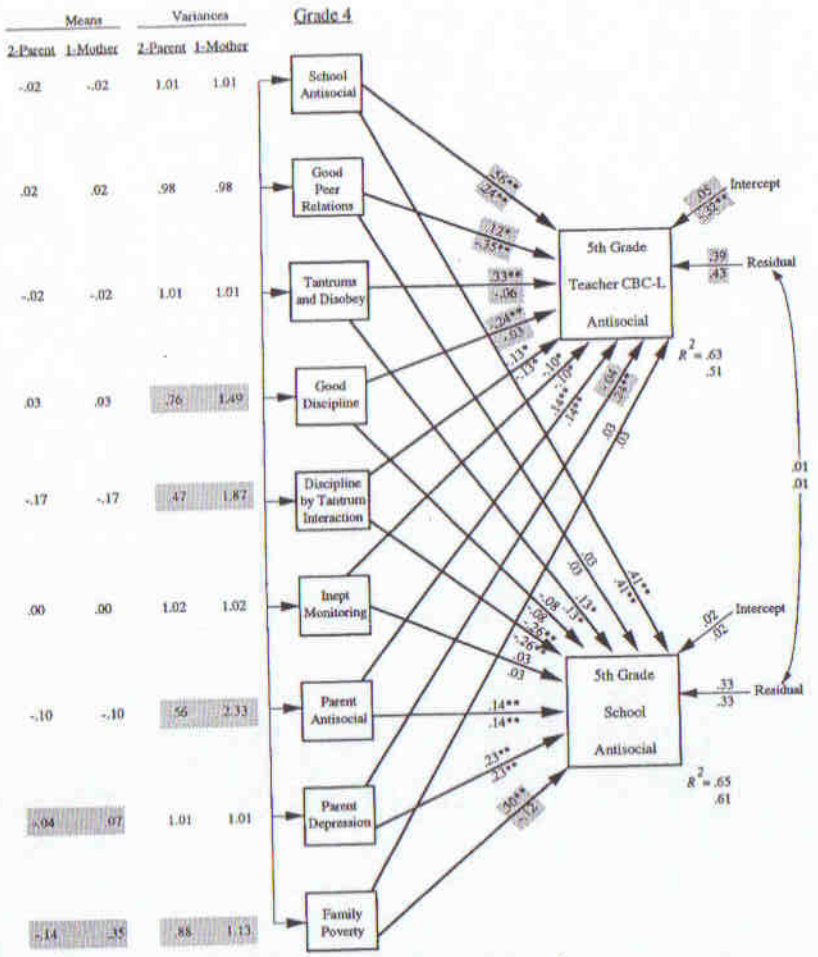

Figure 13.2. Change in School Antisocial Behavior, Partial Constraints Across Family Type

NOTE: Estimates are unstandardized. Two-parent family values shown on top, single-mother values on bottom. Pairs of parameters allowed to be different across family types are indicated by shaded boxes. Two additional pairs of parameters not shown in the diagram, covariances, were allowed to be different across family types.
*p < .05, **p < .01.

opposite effects in the two family-type groups. The effect was negative and strongly significant in the single-mother families but positive and marginally significant in the two-parent families. In contrast, the effect was essentially zero in both family types for the general school antisocial behavior measure. Poverty was the only variable that had significantly different effects across family groups for general school antisocial behavior. Poverty was positive and significant for two-parent families and nonsignificant for single-mother families.

Main effects of key coercion theory variables were variable, but the interaction of tantrums and disobedience with good discipline was significant for both outcomes across both family types. The interaction indicates that the effect of good discipline depends on the level of tantrums and disobedience, and vice versa. The main effects represent the effect of good discipline when tantrums and disobedience is zero, and vice versa. More specifically, because the means of good discipline and tantrums and disobedience are essentially zero in both family-type groups, the main effects represent the effect of good discipline at the mean level of tantrums and disobedience, and vice versa. Following Aiken and West (1991), the basic regression equation can be algebraically manipulated to clarify the dependence of the good discipline effect on the level of tantrums and disobedience. Omitting the other predictors for clarity, the equation becomes

$$(B_{\text{Disc}} + B_{\text{Disc by Tantrums}}\text{Tantrums})\text{Discipline} = \text{School Antisocial}. \quad (13.1)$$

The terms in parentheses represent the simple slope or the effect of discipline at some specified level of tantrums and disobedience. For example, for a tantrums and disobedience score of zero (i.e., the mean level), the simple slope is just the estimated main effect of good discipline, B_{Disc}.

Figure 13.2 shows that the main effects of good discipline and tantrums and disobedience are significant only for two-parent families on teacher ratings of school antisocial behavior. In other words, good discipline is an important predictor of change in teacher ratings of school antisocial behavior at average levels of tantrums and disobedience only for two-parent families. For the other three models, good discipline is important only for above-average levels of tantrums and disobedience. The bulk of the findings in these analyses indicate that

changes in school antisocial behavior are primarily related to the cooccurrence of both a highly coercive child and a lack of good discipline practices. The other key coercion variable, inept monitoring, had small and variable effects across outcomes.

In addition to the key coercion theory variables, both depression and antisocial had significant effects on changes in school antisocial behavior. The effect of parent depression was more variable than was the effect of parent antisocial, being significant for single mothers but nonsignificant for two-parent families on teacher ratings of school antisocial behavior.

In summary, changes in both measures of school antisocial behavior from Grade 4 to Grade 5 are significantly predicted by the interaction of tantrums and disobedience with good discipline for both single-mother families and two-parent families. The cooccurrence of both the lack of skilled parental discipline and a coercive child is at the heart of coercion theory. In addition, measures of parent depression and antisocial had significant effects.

Discussion

Structural equation modeling techniques were used to test a theoretical model for residual change in boys' antisocial behavior at school from Grade 4 to Grade 5. The theoretical model stipulated that family management practices such as skilled parental discipline should be the most proximal predictors of changes in antisocial behavior, whereas parent characteristics such as antisocial tendencies and depression would not contribute once family management practices were included in the model. In addition, other characteristics of the child such as depressed mood, peer relations, and self-esteem, and other family characteristics such as income and socioeconomic status, were tested in preliminary models. Teacher ratings of problematic behavior were collected for the entire sample, and direct observation of classroom and playground behavior, discipline contacts, and additional teacher ratings, combined into an additional antisocial outcome measure, were obtained for a subset of the complete sample. A multisample regression model was estimated across both single-mother families and two-parent families for both Grade 5 school antisocial outcome measures, simultaneously controlling for Grade 4 school antisocial using SEM techniques for missing data.

COERCION THEORY AND CHANGE IN ANTISOCIAL BEHAVIOR

Following previous work by Patterson and his colleagues (i.e., Patterson & Bank, 1989; Patterson, Bank, & Stoolmiller, 1990), it was hypothesized that one of the significant contributors to the preadolescent's antisocial behavior would be the contribution made by the preadolescent himself or herself to the disruption of appropriate parenting practices. Findings from the present study appear to support the contention that coercive behavior of the child, in conjunction with unskilled and coercive parental discipline practices, adequately describe a process that may, in fact, be one of the most significant contributors to the developmental change of the child's antisocial behavior. That is, boys who exhibited higher levels of coercive behavior at home and whose parents demonstrated a lack of appropriate discipline practices were more at risk for experiencing an increase in their antisocial behavior in the school setting.

Parental monitoring practices, however, appeared to be unrelated to changes in antisocial behavior at school from Grade 4 to Grade 5. In other work on the OYS sample, lax supervision has been found to be both an important outcome and predictor of antisocial behavior for boys during middle adolescence (Forgatch & Stoolmiller, 1994; Patterson et al., 1990). Thus it may be that parental discipline practices are more important in childhood whereas supervision practices become more important in adolescence.

Another facet of the present investigation was concerned with examining competing predictors that have been implicated in the formation of adolescent antisocial behavior. Results indicated that parental depression and antisocial behavior were those contextual variables contributing most to the child's development of antisocial behavior between the 4th and 5th grades, effects that were evident for both family types and for both outcome measures. In addition, although restricted to the teacher report of school-based antisocial behavior, findings also support previous work that has demonstrated inadequate peer relations to be a factor in the at-risk adolescent's subsequent submersion into the coercion process. However, results from the present study appear consistent with our position that even when specific contextual and personal variables are accounted for, key coercion theory predictions are still tenable.

It is increasingly clear that antisocial behavior begins early in childhood and is highly stable across development (Loeber & Dishion, 1983). The linchpin of our theoretical model is that the antisocial child's constant conflict with others at both individual and institutional levels is at the core of the youngster's development or continuation of antisocial behavior in each new setting and with each new relationship. Children who have developed an early coercive pattern of interacting with parents will quickly involve themselves in coercive cycles in new settings that will sustain and exacerbate their development toward antisocial behaviors. This demonstrable developmental continuity makes antisocial children acutely at risk for a panoply of other problems such as substance abuse, chronic unemployment, divorce, a range of physical and psychiatric disorders, and dependence on welfare services (Caspi, Elder, & Bem, 1987; Robins & Ratcliff, 1979).

Not only is antisocial behavior highly stable, costly to society, and disruptive to the long-term adjustment of the child, it also is extremely difficult to change. In reviewing the intervention research with antisocial behaviors, Kazdin (1987) concluded that the most promising strategies included a focus on family-management practices, broad based interventions that targeted the child's adjustment in the school, and community interventions that emphasized changing contextual influences on the determinants of antisocial behaviors. Nonetheless, research clearly indicates that as the age-at-referral of the child increases, the probability of successfully intervening and preventing the continued development of antisocial behavior decreases (Dishion, 1984).

ANALYZING INCOMPLETE LONGITUDINAL DATA

Despite the best efforts of researchers, few studies interested in the etiology of antisocial behaviors are completed without the loss of subjects over time or the collection of incomplete data. Data gathered from such studies are most often analyzed using traditional methods such as listwise deletion of missing cases; however, loss of data through attrition, unforeseen budget problems, or other missing data mechanisms can sometimes cause the sample of subjects with complete data to differ substantially from the original sample (Brown, 1990).

Research has shown that traditional estimation methods are inferior to likelihood estimators that use all available information in terms of large sample bias and efficiency (Muthén et al., 1987). The present study involved a situation in which the data could not be considered missing completely at random. However, utilizing SEM techniques, correct maximum-likelihood estimation for the model parameters was obtained.

Such analytical techniques are not, however, without limitations. One problem with the computational approach utilized in the present study is that in many situations there may be a number of missing data patterns so that the number of observations in each group frequently falls below the number of variables. When this occurs it may be necessary to delete some parts of the data, much like the approach of listwise deletion, so that a reasonable number of groups with large sample sizes remain (Muthén et al., 1987). Unless this loss of data introduces strong selective missingness not predictable by observed variables for which there is no missingness, little bias is likely to result. Regardless of the mechanism for the pattern of missingness, utilizing information from a few substantial missing data pattern groups will likely yield more appropriate results than the use of data from only the complete-data group. However, the usual caution concerning generalizability is necessary in that results may be limited not only by the regression model and the pattern of missingness studied but also by the particular specification of the missing data mechanism.

The method described here should make collecting and analyzing longitudinal data in which missing data are likely to occur more attractive to researchers. By combining data from a complete-case subsample and nonresponse subsample, efficient estimates of the coefficients and consistent estimates of their standard errors for a linear model are possible. The method does so introducing only mild assumptions concerning the missing data mechanism beyond those generally made for structural equation modeling.

Summary and Conclusions

In summary, these preliminary analyses of the effect of the key coercion theory variables on the ongoing process of antisocial behavior acknowledge the necessity for differentiating between those developmentally threatening events (e.g., the disruption of family management skills) and those pathogenic processes that are proximal to the development of behavioral problems in the child. Although studies conducted over the past 20 years have demonstrated the ability to predict the likelihood of subsequent delinquency with relative accuracy (e.g., Loeber & Dishion, 1983, 1987), our understanding of the processes by which some children start early and persist into adolescence lags far behind. Of interest, therefore, is whether the present findings add

sufficiently to our understanding of this process to facilitate the development of intervention strategies designed to interrupt the progression of antisocial behaviors.

There is an obvious need for additional research that integrates our basic knowledge of how key coercion variables exert their influence on the ongoing development of serious antisocial behavior, with the assumption that different combinations of these factors are set into motion for different subpopulations of children exposed to major risk factors (Reid, 1991). This assumption becomes increasingly important to efforts aimed at interrupting the coercion process, because it is likely that an increased understanding of the role these mediators play in the etiology of antisocial behavior will ultimately lead to the development of distinct interventions tailored to the needs of the individual.