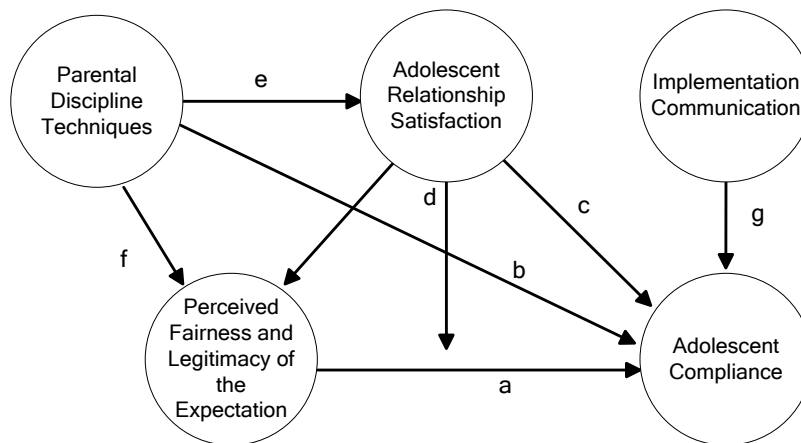


**Guidelines for Writing a Proposal that Uses SEM and for
Writing the Results Section of a Thesis**

Guidelines for Proposal Preparation Using SEM

This handout provides guidelines for writing a dissertation proposal that uses SEM methods. I begin by considering different ways of organizing the introduction section and then I talk about sections that need to be included in the “Analysis” section of the proposal. A separate section presented later describes how to organize the results section after the data have been collected and analyzed, i.e., how to write up the results section for the finished thesis.

I will use as an example the path model below. I want to develop my proposal around a test of this model. The main outcome variable is adolescent compliance with parental expectations about how they are to behave in dating situations. Compliance is impacted by how fair and legitimate the adolescent perceives the expectations as being (path a), the discipline strategies that parents use if a transgression were to occur (path b), how satisfied the adolescent is with his or her relationship with the parents (path c), and implementation communication (path g). Implementation communication refers to how much the parent has talked with the child about how to react when pressures not to behave properly are exerted on them by peers. The other paths in the model are self explanatory (at least for our purposes here).



Writing The Introduction Section

1. When writing the introduction, one strategy is to present your path diagram early as an organizing device that will guide your literature review. The diagram gives the reader an overview of where you are headed and where you will end up. An alternative strategy is to save the presentation of the diagram for the end of the introduction. The idea here is that you review all the literature relevant to the diagram (without ever presenting the diagram) and this review culminates in a synthesized framework that is captured in the diagram. So, the literature review builds up to the diagram. Either approach is acceptable and your choice of which to pursue depends on what you think will communicate best with your readers.

2. Based on the above diagram, I might start my proposal by discussing adolescent compliance with parental expectations (the main outcome variable and topic of interest) and why it is an important area of study. I would then write a literature review that focuses on each link in the theory, with a subheading for each one. For example, I might have a section titled “The Relationship Between Adolescent Compliance and Perceived Fairness/Legitimacy of Expectations,” that corresponds to path a. I might have another section titled “Adolescent Compliance and Adolescent Relationship Satisfaction” that corresponds to path c. And so on.

Within a section, I would review relevant literature and develop the logic of why I think there is a link between the constructs in the heading. I would discuss competing predictions within a section, if they exist. In some cases, it may be natural to consider more than one link within the same section.

3. In doing the above, you will want to keep the measurement model out of the diagram. The introduction is conceptual in nature, so issues of measurement should not be a consideration. The exception, of course, is if your main focus is on measurement issues. Use circles not rectangles in your diagrams. This will make it easier when you introduce latent variables and indicators of those variables in your Method and Results sections.

4. Path diagrams are essentially a set of hypotheses, with one hypothesis per path. Consider ending each subheading with a formal statement of a hypothesis describing the link in the path diagram. For example, I might culminate my discussion in the section on “Discipline Strategies and Adolescent Compliance” with the following hypothesis:

Hypothesis 1: The type of discipline strategy a parent uses will be associated with adolescent compliance with parental behavioral expectations. The more the parent relies on reasoning strategies in conjunction with restricting privileges related to peer contact, the more the adolescent will tend to comply with the expectation.

Note in describing this hypothesis, I avoid causal terms. Many committee members get upset if you use causal terminology because you can never demonstrate causality.

4a. As a variant on this, consider grouping several hypotheses: For example, in my section on “Perceived Legitimacy/Fairness and Adolescent Compliance,” I might end with two hypotheses:

Hypothesis 2: The more legitimate an expectation is perceived as being, the more likely the adolescent will comply with it

Hypothesis 2a: The impact of perceived legitimacy on adolescent compliance is moderated by adolescent relationship satisfaction. The less satisfied an adolescent is, the weaker will be the impact of legitimacy on adolescent compliance. This leads to the prediction that there will be an interaction between perceived legitimacy and relationship satisfaction when predicting compliance. [*Note that it is difficult to avoid causal terminology with moderated relationships*]

4b. As another variant, consider stating mediated relationships as a hypothesis

Hypotheses 1a: The effects of reasoning during discipline on adolescent compliance will be partially mediated by relationship satisfaction and perceived legitimacy. However, reasoning will be associated with compliance independent of these two mediators.

5. If you have more than one viable model, either (1) present a “working model” and then discuss alternatives to it as you discuss the different theoretical links, or (2) present the different models at the outset in different diagrams. I usually find the first strategy works best. Models typically differ in (1) the predicted sign of a path coefficient (one theory predicts a positive relationship, another theory predicts a negative relationship), (2) the presence/absence of a path (one theory says a path exists whereas another theory says it does not), and/or (3) the presence of reciprocal causality (one theory says the causal relationship is reciprocal but the other model says it is unidirectional. How you work these in will depend on the number of differences, how important those differences are, and the complexity of the models.

5a. Don't shy away from formally stating competing hypotheses. For example, in my section on "Parental Discipline Techniques and Adolescent Compliance" I might develop logic for and state a hypothesis that parental use of threats will increase compliance. Then I might turn around and develop logic for why parental use of threats might actually reduce compliance (because the adolescent will rebel and try to get away with everything s/he can). After stating the two hypotheses, I might culminate the section by saying something like:

In sum there are two competing hypotheses that are logically reasonable:

Hypothesis 1a: The greater the number of friends an adolescent has, the less likely they will be to experiment with marijuana.

Hypothesis 1b: The greater the number of friends an adolescent has, the more likely they will be to experiment with marijuana.

The present study will test these competing hypotheses.

Writing the Method Section

In the method section, you will want to present a description of the proposed sample, the way the data will be collected and the measures that you will obtain. You can organize the description of measures around each box/circle in the path diagram. At some point, you will need to talk about how you are going to analyze the data. In some proposals, students create a separate "Results" section to do this and describe how they are going to analyze their data there. In other proposals, it is part of the method section. Use whatever approach your major professor advises you to do and/or what you think works best..

For SEM analyses, after stating what software you will use (e.g., AMOS, M Plus) there are 12 topics you should consider discussing in your data analysis section. You will not necessarily discuss all of these. It depends if they are relevant. The topics are

1. Discuss how you will handle missing data:
2. Discuss how you will handle outliers.
3. Discuss how you will handle non-normality.
4. Discuss model fit indices to be used
5. Discuss limited Information versus full information estimation, if applicable
6. Discuss sample size issues, i.e., statistical power, stability of the sample covariance matrix, and asymptotic theory
7. Discuss measurement error and how it will be dealt with
8. Discuss control of familywise error rates and use of the modified Bonferroni method
9. Discuss factor structure analyses of multiple item measures
10. Discuss clustering and weights, if applicable
11. Discuss model comparison strategies, if applicable (nested chi square strategies)
12. Acknowledge the possibility of redundant/equivalent models

On the next pages, I have written a sample section on each of these topics. Feel free to cut, paste and adapt these to your own proposal. I have no problem with you taking them verbatim, but your advisor may want you to paraphrase. You will need to make changes in a few places based on your sample size and model, but it should be obvious where this needs to be done.

Missing Data

Unless you anticipate having considerable non-normality, you are probably best off using FIML based approaches. If you are going to use FIML, then write the following or some variant of it (I consider the case of attrition in longitudinal designs shortly):

Missing data are expected to be minimal for most variables. Given missing data, parameter estimates and model tests will be pursued in the context of Full Information Maximum Likelihood (FIML) methods as implemented in AMOS/M Plus. Missing data bias will be assessed by computing a dummy variable reflecting the presence or absence of missing data for each variable in the model and then this dummy variable will be correlated with key variables (e.g., demographic variables) in the data.

If you are going to use imputation based approach, then write the following:

Missing data are expected to be minimal for most variables. Where missing values occur, values will be imputed using the Bayesian imputation method in AMOS. If missing data are substantial, a multiple imputation approach will be used with five imputation data sets. Parameters estimates and standard errors across the imputed data sets will be estimated using the formulas in King et al. (2001). Missing data bias will be assessed by computing a dummy variable reflecting the presence or absence of missing data for each variable in the model and then this dummy variable will be correlated with all other variables in the model as well as an array of demographic variables.

(note: if you are using Amelia instead of Bayesian imputation, indicate that this is the case)

If you have a longitudinal design and expect non-trivial attrition, you might add the following in addition to the above:

Although we expect minimal missing data within a wave, we also expect attrition across waves, perhaps as high as 15% to 20%. Missing data will be handled as above. Tests for bias will be supplemented by creating missing value dummy variables at wave t and examining the association between these and scores at time $t-1$.

Outliers

Outlier analyses will be undertaken prior to all major analyses. The analyses will be both non-model based and model based. For the former, multivariate outliers will be identified by examining leverage indices for each individual and defining an outlier as a leverage score four times greater than the mean leverage. If outliers are found, they will be checked for coding errors and the analysis will be conducted both with and without the outliers. Hopefully, results will be comparable across the two forms of analysis. If results differ, then the outliers are consequential and outlier resistant analytic strategies will be considered (Wilcox, 1997, 1999, 2003). An additional set of outlier analyses will be pursued using model-based outlier analysis. This involves selecting an indicator for each latent variable (or, for the case of single indicators, using the only indicator available) and then regressing the indicator for each endogenous variable onto the indicators for variables that the endogenous variable is assumed to be a linear function of. This analysis will use ordinary least squares regression in a limited information estimation framework. Standardized $dfbetas$ will be examined for each individual and each predictor as well as the intercept. An outlier will be defined as anyone with an absolute standardized $dfbeta$ larger than 1.0. If outliers are observed, the analysis will be conducted both

with and without the outliers to evaluate their consequentiality. If necessary, outlier resistant estimation will be pursued (Wilcox, 1997, 2003).

Non-normality

If you are going to use bootstrapping should you encounter non-normality, write the following:

Multivariate normality will be evaluated using Mardia's test for multivariate normality. In addition, univariate indices of skewness and kurtosis will be examined to determine if the absolute value of any of these indices is greater than 2.0. If non-normality appears to be problematic, then bootstrapping will be pursued as a remedy. P values and confidence intervals will be estimated using bias-corrected methods. The number of bootstrap replicates will be 2000. In place of the traditional chi square test, the Bollen-Stine bootstrapped version of the test will be performed.

If you are going to use robust maximum likelihood should you encounter non-normality, write the following:

Multivariate normality will be evaluated using Mardia's test for multivariate normality. In addition, univariate indices of skewness and kurtosis will be examined to determine if the absolute value of any of these indices is greater than 2.0. If non-normality appears to be problematic, then robust methods of maximum likelihood analysis will be used, employing the algorithms in M Plus.

Indices of Fit

Following the recommendations of Bollen and Long (1993), a variety of global fit indices will be used, including indices of absolute fit, indices of relative fit and indices of fit with a penalty function for lack of parsimony. These include the traditional overall chi square test of model fit (which should be statistically non-significant), the Root Mean Square Error of Approximation (RMSEA; which should be less than 0.08 to declare satisfactory fit), the p value for the test of close fit (which should be statistically non-significant), the Comparative Fit Index (CFI; which should be greater than 0.95); and the standardized root mean square residual (which should be less than 0.05).

In addition to the global fit indices, more focused tests of fit will be pursued. These include examination of the standardized residual covariances (which should be between -2.00 and 2.00) and modification indices (which should be less than 4.00). The parameter estimates also will be examined for Heywood cases. Care will be taken to ensure there is no specification error.

Limited Information Estimation versus Full Information Estimation

Most thesis proposals will not need to use this section:

The theoretical questions posed in this research are framed in the path diagram in Figure 1. It is natural to think of applying traditional structural equation modeling (SEM) strategies to such models. SEM computer programs use full information estimation approaches where all of the path coefficients (and their standard errors) are estimated simultaneously in the context of the full system of linear equations implied by the model. The same statistical algorithm (e.g., maximum likelihood estimation) is applied throughout. An alternative approach is to use a limited information estimation strategy. This approach uses the path diagram to identify the structural relationships of interest and to define the relevant linear equations. However, the

overall model is broken up into pieces and estimates of the coefficients are derived within each piece separately using statistical methods that are appropriate for that piece. Full information estimation approaches can yield more efficient estimates and also yield more plentiful statistics about goodness of model fit. However, the full information estimation approach also has disadvantages. For example, model misspecification in one part of the model can yield biased estimates in another part of the model. By contrast, in limited information estimation, specification error is compartmentalized. Limited information estimation also allows one to tailor the analytic method to the nature of the variables involved in a given piece of the overall model (e.g., logistic regression, ordinal regression, OLS regression, Poisson regression). Full information estimation strategies will be pursued but, where necessary, limited information estimation approaches will be used.

Sample Size Considerations

To determine an appropriate sample size, structural equation modeling requires that in addition to statistical power, issues of the stability of the covariance matrix and the use of asymptotic theory be taken into account. In terms of power, it is difficult to evaluate the power associated with specific path coefficients in complex SEM models because of the large number of assumptions about population parameters that must be made. A rough approximation of power can be obtained by using a limited information approach with single indicators of the path models implied by Figure 1. This permits the use of traditional power analysis software to gain a sense of sample size demands (Jaccard & Wan, 1996). In all examples below, we assume an alpha level of 0.05 and a two tailed test.

For a multiple regression analysis with 4 predictors where the squared multiple correlation is 0.30 and where one wants to detect a predictor that accounts for at least 5% unique variance in the outcome, the required sample size to achieve power of 0.80 is approximately 115. For a logistic regression analysis where the target predictor is a continuous predictor with four other predictors in the equation, where the event rate at the mean of all predictors is 0.20 and where the multiple correlation of the predictor with the other predictors is 0.30, the sample size needed to detect an odds ratio of 1.75 expressed in standardized metrics is about 170 and for an odds ratio in standardized metrics of 2.00 is about 110. For a simple zero order correlation of 0.30 in the population, the sample size needed to achieve power of 0.80 is approximately 80. For a contrast of means between two independent groups and an effect size corresponding to Cohen's definition of a medium effect (a d value of 0.50), the sample size needed to achieve power of 0.80 is approximately 65 per group. For a contrast of dependent means, the corresponding required sample size is about 35. For a percentage difference between two independent groups where the population percentage in the first group is 30 percent and in the second group it is 15 percent, the required sample size for power of 0.80 is about 120 per group. The proposed sample size for this study seems adequate in terms of power.

[Note: You will omit or include some of the above statements depending on what tests you are doing, and you might approach the description of sample sizes a bit differently, as discussed in class].

In terms of asymptotic theory and covariance stability, simulation studies tend to suggest that sample sizes of 100 to 125 or larger often yield adequate results given that reasonably reliable measures are used (reliabilities greater than 0.65) and with a reasonable number of indicators per latent variable (Jackson, 2003; Jaccard & Wan, 1996). The sample size in the proposed study exceeds this standard.

[If your sample size is less than 100 or 125 then consider adding the following]: Because the proposed sample size is somewhat less than this, two small scale simulation studies were conducted to determine if the proposed sample size is reasonable when focusing on selected portions of the theoretical framework in Figure 1. One simulation study evaluated the behavior of the traditional chi square index of model fit and a test of a correlation between factors in a two factor confirmatory factor analysis model with three indicators per factor. In this simulation, the reliability of each indicator was 0.70 and the correlation between factors was set to zero to evaluate the Type I error rate. The model has 8 degrees of freedom. A theoretical chi square distribution with 8 degrees of freedom has a mean of 8 and a standard deviation of 4. The simulation used 5000 replicates each with a sample size of 75 **[note: you will use whatever sample size you are proposing here]**. All 5000 solutions converged. Across the 5000 replications, the mean chi square was 8.5 and the standard deviation was 4.2. Examination of centiles spanning the observed and theoretical distributions indicated close correspondence between the distributions, even at the extreme ends where significance conclusions are made. The Type I error rate for the factor correlation was 0.068 and the coverage for the 95% confidence interval was 93.2%. When the factor correlation was set to .33, the 95% confidence interval coverage was 93%.

A second simulation used a latent variable multiple regression analysis with a sample size of 75. A latent Y variable was regressed onto four latent predictors, each with three indicators that had reliabilities of 0.70. The three latent predictors were all assumed to be correlated 0.30 with one another (a commonly observed correlation between variables reported in the literature – see Maxwell, 2002) and the overall squared multiple correlation was set to 0.30. To evaluate the chi square index of fit and Type I error rates, one of the predictor regression coefficients was set to zero while the others were defined so that each had a standardized regression coefficient of 0.25. The model has 81 degrees of freedom. A theoretical chi square distribution with 81 degrees of freedom has a mean of 81 and a standard deviation of 12.7. The simulation used 5000 replicates each with a sample size of 75. All 5000 solutions converged. Across the 5000 replications, the mean chi square was 91.3 and the standard deviation was 14.2. Examination of centiles spanning the observed and theoretical distributions indicated reasonably close correspondence between the distributions, but at the extreme ends of the distribution where significance conclusions are made, there was a tendency for the observed chi square to be somewhat larger than expected. However, the type I error rate for the zero regression coefficient was 0.059 and the coverage for the 95% confidence interval was 94.1%. For the non-zero regression coefficients, the 95% confidence interval coverage was approximately 93%.

Overall, these results suggest that the use of asymptotic theory will be acceptable given the proposed sample size and that the covariance matrices should be sufficiently stable. Although the final model to be evaluated will be more complex, the simulation studies suggest that the proposed sample size probably is viable.

Measurement Error

Measurement error will be taken into account through the use of multiple indicators of constructs. In cases where only a single indicator is available, we will adopt the strategy suggested by Joreskog and Sorbom (1996). This involves constraining the error/unique variances for each measure to values corresponding to a priori determined levels of reliability. The reliability levels for the measures will be based on alpha coefficients or previous research. ***[Describe the values that will be used and the rationale for each]***

Familywise Error Rates and Multiple Contrasts

At times, multiple significance tests will be conducted within a family of contrasts and there will be concern for inflated experimentwise error rates. The robustness of our conclusions will be compared both with and without statistical corrections for multiple tests (using the strategy discussed in Jaccard & Guilamo-Ramos, 2002). In general, we will use a Holm adjusted modified Bonferroni method (Jaccard, 1998) for controlling experimentwise error rates, which is more powerful than traditional Bonferroni methods.

Factor Structure of Multiple Items Measures

For all multi-item measures, the coefficient alphas and factor structures of the measures will be evaluated to ensure that they are behaving in a way that one would expect based on their psychometric histories. Some of the variables in the path diagrams reflect variable categories with multiple variables or dimensions. The intercorrelations of variables will routinely be examined, and coupled with substantive criteria and the results of exploratory or confirmatory factor analyses, decisions will be made about combining indices or introducing latent constructs into the analysis.

Sample Weights

Most proposals will not need this section. However, if you have sample weights, consider adding the following:

The use of sampling weights in complex model evaluation is controversial (e.g., Lohr & Liu, 1994; Winship & Radbill, 1994). Winship and Radbill (1994) note that if a model is specified correctly and sampling is not outcome based, then use of unweighted estimation strategies are preferred over weighted estimation strategies because they yield smaller standard errors. In practice, models are almost always misspecified to some degree, so the more realistic question is whether the degree of misspecification is consequential (Kott, 1991; DuMouchel & Duncan, 1983). Feinberg (1989) argues that outcome based sampling is the only situation in which weights should be used for multivariate analyses. A range of perspectives on the use of sample weights can be found in Saphire (1984), Rubin (1985), Little (1991), Lohr and Liu (1994), Winship and Randall (1994) and Scott and Wild (1989). DuMouchel and Duncan (1983) propose a test of whether the results of a weighted solution differs significantly from those of unweighted solutions. Asparouhov and Muthen (2009) present a comparable test for SEM modeling. In the absence of such differences, one can report either weighted or unweighted results. The results of the weighted analyses are reported if one wants to be conservative with respect to bias and specification error whereas the results of the unweighted analyses are reported if one wants to maximize efficiency of the estimators.

Both unweighted and weighted analyses will be pursued. The weighted analyses will be performed in M Plus using the procedures discussed in Asparouhov (2005).

Clustering

The data will be collected in different organizations/schools with a substantial number of persons within each organization/school, so there is the possibility of clustering effects. We will evaluate this by examining intraclass correlations and adjust for clustering if it seems necessary to do so, either by the introduction of covariates reflecting organization/school units, or the use of robust estimators available in the M Plus computer programs.

Model Comparisons

Comparisons of nested models will use the traditional nested chi square test. For cases where non-normality is present, a robust nested chi square test will be used as implemented in M Plus (2004). For non-nested models, comparisons will be made using the Bayesian Information Criterion (Raftery, 1995).

Equivalent Models

It is recognized that there may be equivalent models that can account for the data relative to the models being tested. Equivalent models will be described and used to qualify conclusions in the discussion section.

References

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, 12, 411-434.
- Bollen, K. & Long, S. (1993). *Testing structural equation models*. Newbury Park: Sage.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Matwah, NJ: Erlbaum.
- DuMouchel, W.H. & Duncan, G. J. (1983). Using sampling survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78, 535-543.
- Feinberg, S. E. (1989). Modeling considerations: Discussion from a modeling perspective. In Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P. (Eds.) *Panel surveys*. New York: Wiley.
- Honaker, J., Joseph, A., King, G., Scheve, K. & Singh, N. (2003). *Amelia: A program for missing data*. Department of Government, Harvard University.
- Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Newbury Park: Sage.
- Jaccard, J. and Wan, C. (1996). *LISREL analyses of interaction effects in multiple regression*. Newbury Park: Sage.
- Jaccard, J. (2006). *Multiple regression analysis in clinical child and adolescent psychology*. *Journal of Clinical Child and Adolescent Psychology*, In press.
- Jaccard, J. and Guilamo-Ramos (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Basic issues and recommendations. *Journal of Clinical Child and Adolescent Psychology*, 31, 130-146.
- Jackson, D. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, 10, 128–141.
- Joreskog, K. & Sorbom, D, (1996). *Users manual for LISREL*. Chicago: Scientific Software.
- King, G., Honaker, J., Joseph, A. & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95, 49-69.
- Kott, P.S. (1991). A model based look at linear regression with survey data. *American Statistician*, 45, 107-112.
- Liang, K. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Little, R.J.A. (1991). Inference with survey weights. *Journal of Official Statistics*, 7, 405-424.
- Lohr, S. L & Liu, J. (1994). A comparison of weighted and unweighted analyses in the National Crime Victimization Survey. *Journal of Quantitative Criminology*, 10, 343-360.
- Maxwell, S. (2000). Sample size and multiple regression analysis. *Psychological Methods*, 5, 434-458.

- Muthén, B.O. (2004). Mplus technical appendices. Los Angeles, CA: Muthén & Muthén
- Muthen, L. & Muthen, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural-Equation-Modeling*, 9, 599-620. (Power1.pdf)
- Raftery, A. (1995). Bayesian model selection in social science research. *Sociological Methodology*, 25, 111-163.
- Rubin, D.B. (1985). The use of propensity scores in applied Bayesian inference. In Bernardo, J., DeGroot, M.H., Lindley, D.V. & Smith, A. (Eds.). *Bayesian statistics*. Amsterdam: North Holland.
- Saphire, D.G. (1984). *Estimation of victimization prevalence using data from the National Crime Survey*. New York: Springer-Verlag.
- Scott, A.J. & Wild, C.J. (1989). Selection based on the response variable in logistic regression. In Skinner, C.J., Holt, D. & Smith, T.M. (Eds.). *Analysis of complex surveys*. New York: Wiley.
- Wilcox, R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press.
- Wilcox, R. (1999). *Fundamentals of modern statistical methods*. New York: Springer.
- Wilcox, R. (2003). *Applying contemporary statistical techniques*. San Diego: Academic Press.
- Winship, C. & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods and Research*, 23, 230-257.

Guidelines for Writing a Results Section in a Thesis

This section provides guidelines for writing the results section of a thesis once data analysis is complete. You will want to cover the same basic sections as your proposal, but now the description of what you did will be integrated with the presentation of results.

A common approach is to have three sections, each labeled with a separate heading, (1) preliminary analyses, (2) main analyses, and (3) supplementary analyses.

Preliminary Analyses

In this section, you describe how much missing data occurred, what biases were isolated in missing data patterns (if any), how much non-normality there was and what happened in your analysis of outliers. In addition you describe any psychometric analyses that were done on your measures.

Here is a section titled “Preliminary Analyses” from an article I published that addresses most of these issues and illustrates this portion of a results sections.

Preliminary Analyses

Descriptive Statistics. Table 1 presents means and standard deviations for all of the continuous variables used in the models. The median values for each of the variables (not reported) were close to the mean values. Mean scores for MPQ-SF sensory and affective indices were within average levels observed in other studies (Melzack, 1987). The mean value of T-STAI corresponds to T scores of 55 (female) and 56 (male) for normal adults age 40-49.

Outliers. Both model based and non-model based outlier analyses were pursued. For the former, a leverage score was calculated for each respondent based on their multivariate profile for the nine variables included in model analyses. The mean leverage score across respondents was .054 and an outlier was defined as anyone having a leverage score three times the value of the mean (Jaccard & Wan, 2003). No outliers were evident using this criterion. Model based outliers were examined using limited information regression analyses for each of the linear equations dictated by the various path models tested (Bollen, 1996). We examined DfBeta values for each individual relative to each path coefficient to isolate unusually influential individuals in parameter estimation. An outlier was defined as individuals who had dfBetas three times larger than the standard error of a coefficient. No outliers were evident in these analyses.

Missing Data. There were small amounts of missing data amounting to no more than a few cases on any given variable. There was no coherent pattern to the missing data. For those individuals with missing data, values were imputed to conform to covariance estimates consistent with the application of the Bayesian algorithm implemented in AMOS (Schafer, 1997). Given the small number of instances of missing data, concerns surrounding estimation with missing information are moot.

Non-Normality. Traditional maximum likelihood methods of SEM assume that the continuous variables in the model are multivariately normally distributed. This was tested using the Mardia test, which yielded a statistically significant result (critical ratio = 2.56, $p < 0.05$). This suggests the presence of non-normality at the multivariate level. Skewness and kurtosis indices for each variable are presented in Table 1. Troublesome skewness and kurtosis values are evident for the measure of psychopathology. Given this, the decision was made to pursue

parameter estimation using bootstrapping. We performed 2,000 bootstrap replications for purposes of estimating standard errors, p values, and confidence intervals. We used the bias corrected approach to interval estimation as implemented in the computer program AMOS. Estimation for the individual bootstrap samples was well-behaved and yielded convergence and meaningful solutions in all 2,000 instances. The p value for overall fit of the tested models was calculated using the Bollen-Stine bootstrap approach in place of the traditional chi square statistic (Bollen & Stine, 1993). In general, conclusions were the same in both estimation approaches. All significance tests and confidence intervals reported are from the bootstrap analyses.

Main Analyses

In this section you want to convey what happened in your primary analyses, following the basic write-ups we used in class. If your initial model did not fit and you made modifications to it, you will want to note this, as discussed in class. You do not want to take the reader through all the gory details, but you do want to let them know how the analysis progressed from beginning to end. Here is an example write up from a paper I wrote on a project I consulted on. The introduction to the paper had set up the model in Figure 1, reviewing all literature relevant to it and making a case that it was novel and had something to contribute. In the sections below, you will see that after laying out the statistics, I add paragraphs that explicitly discuss some of the more important paths,.

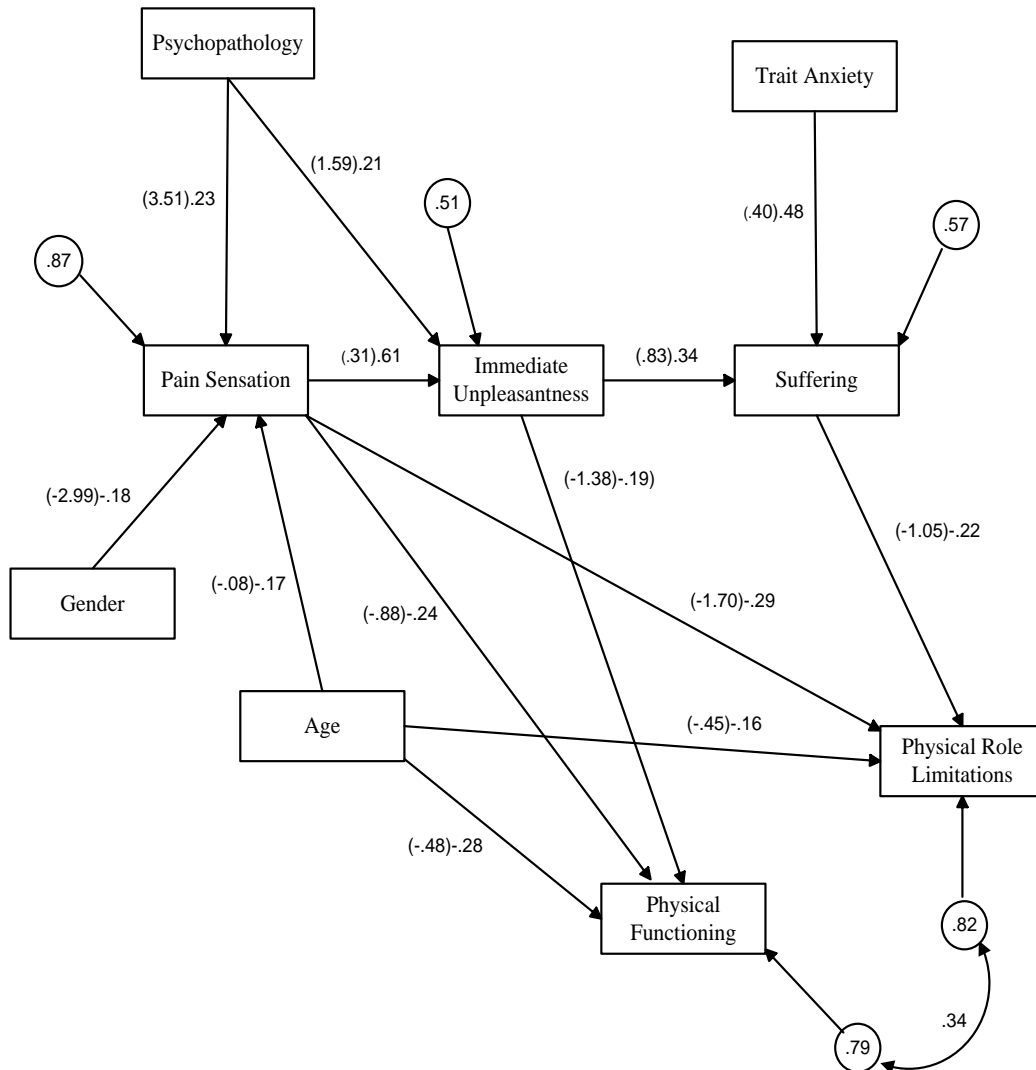
Model Tests

The initial model tested can be described in conjunction with Figure 1, which represents the final model that we settled upon. The central feature of the model was the sequential causal links between pain sensation, pain affect, long-term pain suffering and behavior. In this case, there were two behavioral outcomes, physical role limitations and physical functioning. Although there is no causal link depicted in Figure 1 between long term suffering and physical functioning, this path was included in the initial model. All other variables in Figure 1 were included in the model and assumed the role of exogenous variables. Causal paths were defined from each exogenous variable to each of the five endogenous variables, creating a saturated model (with the exception of the patterning of causal links between the five endogenous variables). All residual variances (reflected by the circles in Figure 1) were assumed to be uncorrelated and all exogenous variables were assumed to be correlated. The Bollen-Stine p value for this model was statistically significant ($p < 0.001$) suggesting poor model fit. The more traditional indices of global fit yielded a mixed picture ($\chi^2 = 26.76$, $df = 4$, $p < 0.001$; CFI = 0.94, RMSEA = 0.19; close fit test p value < 0.001 ; standardized RMR = 0.045).

Inspection of model diagnostics revealed that the sole source of ill fit was the assumption of uncorrelated residuals for the two behavioral outcomes, physical functioning and physical role limitations. Two strategies were considered for dealing with the ill fit. One strategy was to model the two illness behavior measures as indicators of a common latent variable reflecting a more global construct of physical impairment. This strategy was rejected because diagnostics also made evident that the two variables were differentially related to other constructs in the model. Treating them as indicators of the same underlying construct would obscure these differences. The second strategy was to maintain the conceptual distinctions between the variables but to permit the residuals to be correlated. This is justified theoretically if one can specify variables outside of the theoretical system that might serve as common causes of the two constructs. Because specification of such variables is straightforward (e.g., health status in general, overall physical fitness), we adopted this strategy.¹ The revised model with correlated error was re-fit to the data and the model yielded good fit. The Bollen-Stine p value

was 0.26 and all of the traditional indices of overall fit were satisfactory ($\chi^2 = 5.31$, $df = 3$, $p < 0.15$; CFI = 0.99, RMSEA = <0.07 ; close fit test p value < 0.29 ; standardized RMR = 0.018). In addition, more focused fit tests (examination of modification indices, offending estimates, standardized residuals and evaluations of theoretical coherence) all suggested adequate model fit.

Figure 1: Final SEM Model for Pain Processing Stages



We examined the path coefficients for this model and deleted all paths from the model that were not statistically significant. To control for chance effects across multiple tests of significance, we adopted a modified Bonferroni criterion for declaring statistical significance of a path coefficient based on the False Discovery Rate (FDR) method (Keselman, Cribbie, & Holland, 1999). Using this method, a family of tests was defined as the path coefficients leading from the exogenous variables to a given endogenous variable. As convention dictates, within-family error rates were controlled using the FDR, but controls across families were not invoked. This yielded the model in Figure 1. The trimmed model was re-fit and good model fit was still manifest. The Bollen-Stine p value was 0.26 and the traditional fit indices were $\chi^2 = 21.67$, $df =$

17, $p < 0.20$; GFI = 0.97; CFI = 0.99, RMSEA = <0.04 ; close fit test p value < 0.58 ; standardized RMR = 0.046.

The path coefficients in Figure 1 are from the trimmed model. Both unstandardized and standardized path coefficients are presented, with unstandardized coefficients in parentheses. All residuals and correlations are in standardized metrics. Correlations between the exogenous variables are omitted for purposes of figure clarity. All path coefficients were statistically significant ($p < 0.05$). Table 2 presents the bias corrected confidence intervals from the bootstrap analyses for the unstandardized coefficients. None of the path coefficients in the trimmed model were impacted much by the trimming (i.e., their values in the more saturated model were comparable to those in the trimmed model). All were statistically significant before trimming and all remained statistically significant after trimming. We highlight results in terms of the conceptual questions outlined in the introduction.

Sequential Model of Pain Processing. The sequential model of pain processing predicts that the path coefficients from pain sensation to pain affect, from pain affect to suffering, and from suffering to behavior should be statistically significant. In general, this was the case. However, there were several notable results that were contrary to the model when just the four central constructs of sensation, affect, suffering, and behavior are considered. First, suffering was a statistically significant predictor of only one of the illness behaviors (physical role limitations). Second, pain sensation had statistically significant direct effects on illness behavior independent of pain affect or pain suffering. Whereas previous research has suggested that the effects of pain sensation on behavior should be mediated by affective responses to pain, our results suggest that pain sensation can affect behavior independent of these mechanisms. To be sure, there is evidence that some mediation does, in fact, occur (as reflected by supplemental tests based on the bootstrapped analyses that we performed demonstrating a statistically significant reduction in the path coefficient from pain sensation to physical role limitations when the mediators of pain affect and suffering were held constant; see (Hoyle & Kenny, 1999), for a description of the general logic of such tests). But despite this, the data also suggest that pain sensation affects behavior over and above any effects that it has on these mediators.

The model in Figure 1 is consistent with the proposition that (1) pain affect mediates the effect of pain sensation on pain suffering and (2) that pain suffering mediates the effects of pain affect on behavior (when pain affect is related to behavior). Supplemental analyses on the reduction of path coefficients when the mediators were held constant yielded results consistent with complete mediation. Both of these propositions are in accord with past research on the sequential model of pain processing.

Gender Effects. The statistically significant path coefficient in Figure 1 from gender to pain suggests gender differences in pain sensation (holding age and psychopathology constant). The unstandardized coefficient reflects the adjusted mean difference between males and females and indicates that, on average, females report higher levels of pain sensation than males (by about three scale units). Although gender did not have direct effects on pain affect, suffering, or behavior, the model suggests that this is because the effects of gender on these variables are mediated by pain sensation. When pain sensation is held constant, gender effects on these more distal variables reduces by a statistically significant amount and, in fact, to non-significant levels of prediction. Thus, the key to understanding gender differences on the more distal outcome variables seems to be the differential effects that gender has on pain sensation (holding age and psychopathology constant).

Age. Like gender, there was a statistically significant path coefficient between age and pain sensation, with pain sensation tending to decrease as age increases. Although pain sensation seems to mediate the effects of age on pain affect, pain suffering, and to some extent, behavior, age also had independent effects on behavior over and above these mediators. When the effects of pain sensation were held constant, older individuals tended to report greater problems related to physical health than younger individuals.

Chronic Anxiety. The only statistically significant path coefficient for trait anxiety was from anxiety to pain suffering. In general, higher levels of anxiety were associated with higher levels of suffering (holding constant pain affect). Anxiety also had an indirect effect on physical role limitations (with higher anxious patients exhibiting more limitations and problems), but these effects were mediated by pain suffering. When long-term pain suffering was statistically held constant, the effects of anxiety on physical role limitations were reduced to non-significance.

Psychopathology. Psychopathology yielded two statistically significant path coefficients. The first was to pain sensation, with individuals characterized by higher levels of psychopathology reporting higher levels of pain sensation (holding gender and age constant). Psychopathology also had statistically significant effects on pain affect, some of which was due to the mediating role of pain sensation, i.e. because individuals with higher levels of psychopathology report higher levels of pain sensation and because those with higher levels of pain sensation report higher levels of pain affect, it follows that those with higher levels of psychopathology also report higher levels of pain affect. However, the model in Figure 1 suggests that the effect of psychopathology on pain sensation cannot completely account for the effects of psychopathology on pain affect. Rather, psychopathology seems to have an independent effect on pain affect irrespective of pain sensation. These effects of psychopathology on pain sensation and pain affect work their way through the theoretical system to ultimately produce effects on both pain suffering and behavior. However, the key mechanisms to understanding the impact of psychopathology on pain suffering and behavior seems to be those relating psychopathology to pain sensation and pain affect.

Supplemental Analyses

In this section, you want to address supplemental issues, such as assuring the reader that the study was sufficiently powered, the biasing effects of measurement error, and exploring specification error. Here is an example from the same article I reported the main results from above.

Supplemental Analyses

In addition to the above model tests, we conducted supplementary analyses to explore potential problems of model misspecification and parameter bias induced by the presence of measurement error. With respect to the former, we used traditional regression methods in conjunction with product terms to test for possible interaction effects between predictors of each endogenous variable in Figure 1 (Jaccard & Wan, 2003). The regression equations were dictated by the limited information estimation approach to SEM described by Bollen (1996) and did not suggest the presence of any meaningful interaction effects. In terms of measurement error, we re-estimated the model in Figure 1 but imposed an a priori determined amount of measurement error onto the observed measures using the strategy described by Joreskog and Sorbom (1997). The amount of unreliability imposed was based on the alpha coefficient for each scale (i.e., the proportion of random error due to measurement error was set to be 1 minus the alpha coefficient for the scale). None of the major conclusions drawn from the original significance tests were changed.

It also is useful to provide perspectives on statistical power for the tests of the path coefficients so that one can better appreciate the possibility of a Type II error for statistically non-significant path coefficients. Power analyses for SEM models are complicated and often rest on assumptions that are impractical or not viable. We followed the practice recommended by Jaccard and Turrisi (2003) that provides a rough sense of statistical power by applying power analytic methods for OLS regression as applied to selected linear equations from the set of linear equations implied by the model in question. Given a sample size of 168 and a two tailed alpha level of 0.05, we evaluated the statistical power associated with a path coefficient that represents 5% explained variance over and above a set of five additional covariates. Based on the residuals in Figure 1, we evaluated three scenarios where the initial set of covariates accounted for 10% of the variance, 20% of the variance, or 40% of the variance. The approximate statistical power in these three scenarios was 0.87, 0.89, and 0.97. For a path coefficient that represents 3% additional explained variance in the same scenarios, the approximate statistical power was 0.66, 0.72, and 0.84. Overall, the approximate power seems adequate for detecting paths that account for at least 5% of the variance of an outcome variable and in some cases, it also is adequate for coefficients that reflect only 3% unique explained variance.