Multiple Linear Regression Viewpoints

Table of Contents

Multiple Linear Regression Viewpoints

Comparing OLS and HLM Models and the Questions They Answer: Potential Concerns for Type VI Errors

| **David Newman** | **Isadore Newman** | **James Salzman** |
| Cleveland State University | Florida International University | Ohio University |

Hierarchical Linear Modeling (HLM) has become an important analytical tool in a number of fields of study and many of the educational journal articles published during the last decade have used this technique. In this study, the authors proposed that a clearer understanding of what HLM was truly testing could be facilitated if the researcher constructed multiple linear regression (MLR) models to reflect the research questions that the corresponding HLM was intending to test. The authors compared the results of HLM and MLR models on a sample data set, discussed the advantages and disadvantages of each technique, and provided some examples of the increased risk of committing Type VI errors in running an HLM model without understanding the underlying question that was being asked. Further issues that pertained to researchers' obligations to differentiate between statistical analysis and research design were discussed in the areas of centering, interaction, and fixed, mixed, and randomized effects.

Hierarchical Linear Models (HLM), which are frequently referred to as multilevel models, are most appropriately and effectively used when variables tend to be nested within other variables. For example, individual students may be nested within classes and classes could be nested within schools. A number of researchers (Field, 2009; Kreft, 1996; Morris, 1995; Mundfrom & Schultz, 2002; Raudenbush & Bryk, 2002; Tabachnick & Fidell, 2007) have indicated that HLM is superior to ordinary least squares (OLS/General Linear Models) because HLM theoretically produces appropriate error terms that control for potential dependency due to nesting effects, while OLS does not. An additional argument favoring the use of HLM is that it is a generalization of OLS, which better handles continuous variables that reflect randomized effect designs, and, therefore, HLM produces more accurate error terms and Type I error rates (Mundfrom & Schultz; Raudenbush, 2009; Raudenbush & Bryk). A good part of the cited advantages for HLM is related to the situations in which the intraclass correlations, which is the between group effect divided by the total effect, departs from zero with $\rho = \tau_{00} /(\tau_{00} + \sigma^2)$. If the correlation is zero, there seems to be less advantage to using HLM because there is no interclass correlation.

Method

It is the contention of this paper that the stated advantage for using HLM instead of the traditional OLS/Multiple Linear Regression (MLR) models, is that the MLR models typically were not written to most appropriately reflect the research questions in nested designs. We believe this is frequently due to the failure to incorporate person vectors (McNeil, Newman & Kelly, 1996; Pedazur, 1982; Williams, 1987). A person vector is binary coded so that if a dependent score comes from a particular unit, such as a person who is repeated, a school, or a district, etc., it is coded as 1 and coded as a 0 if it does not come from that unit. The person vector coding, thus, controls for the variability due to the unit. Without this type of coding, there tends to be a confounding or nesting effect that leads to an inconsistency between the research question and designing an appropriate model to test the question. We refer to this inconsistency as a Type VI Error (Newman, Fraas, Newman & Brown, 2002; Tracz, Nelson, Newman & Beltran, 2005; Tracz, Newman, Nelson, & Dellran, 2004).

Results and Discussion

OLS Models

The following examples were run to demonstrate how different models could be written in OLS/MLR to reflect different research questions. For example, if one were interested in determining if a treatment predicts Ohio Achievement Test (OAT) scores, the following three models could be written to test this question:

Table 1. Model 1: $OAT = a_0u + a_1$ (Treatment) + Error

Model Summary

Model					Change Statistics				
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.038[a]	.001	.001	28.791	.001	2.679	1	1819	.102

a. Predictors: (Constant), TX

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	404.790	.954		424.360	.000
	TX	2.209	1.349	.038	1.637	.102

a. Dependent Variable: OAT_09

This model would be an over simplification of the modeling of the question of interest if the treatment were nested within schools. The results of this model in Table 1 indicate that $a_1 = 2.209$ with a $p = .102$; meaning the treatment was not statistically significant at alpha = .05. A more appropriate model to reflect the question of interest when treatment is nested would be:

Model 2: $OAT = b_{01} + b_{37}(School\_1) + b_{38}(School\_2) + b_{39}(School\_3) +$ $b_{69}(School\_34) + E$
Model 3: $OAT = b_0u + b_1$ (Treatment) $+ b_2$ (School 1) $+ \ldots + b_N$ (School N) $+ E$

Table 2. Model 2 & 3 Summary.

Model Summary

Model					Change Statistics				
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
2	.385[a]	.148	.132	26.841	.148	9.117	34	1786	.000
3	.388[b]	.151	.134	26.801	.003	6.376	1	1785	.012

a. Predictors: (Constant), School_34, School_28, School_14, School_23, School_33, School_15, School_10, School_6, School_8, School_30, School_11, School_16, School_32, School_31, School_22, School_24, School_19, School_4, School_5, School_25, School_7, School_12, School_21, School_29, School_2, School_26, School_20, School_13, School_17, School_18, School_9, School_27, School_1, School_3

b. Predictors: (Constant), School_34, School_28, School_14, School_23, School_33, School_15, School_10, School_6, School_8, School_30, School_11, School_16, School_32, School_31, School_22, School_24, School_19, School_4, School_5, School_25, School_7, School_12, School_21, School_29, School_2, School_26, School_20, School_13, School_17, School_18, School_9, School_27, School_1, School_3, TX

The first model is simply testing to see if there is a relationship between people who had the treatment and the OAT scores. Testing the second model against Model 3 is considerably different in that it is testing to see if the treatment has an effect independent of school differences. Please note that b_1 is now 4.016 compared to a_1 in Model 1, which is 2.209. As seen in Table 2, when Model 2 is tested against Model 3, there is a statistically significant difference ($p = .012$). By adding in the school vectors, the model more accurately reflects the question of interest. As one can see, the number of independent vectors has increased by the number of schools minus one (K-1) for both the full and restricted models. Table 3 shows that when the random nature and violation of independence of errors are accounted for, the fixed effect for treatment now becomes statistically significant with $p = .012$ and a regression coefficient = 4.016.

Table 3. Models 2 & 3 Coefficients Table

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
2	(Constant)	415.638	3.524		117.932	0
	School_1	0.798	4.224	0.007	0.189	0.85
	School_2	-7.527	4.884	-0.048	-1.541	0.123
	School_3	6.033	4.121	0.059	1.464	0.143
	.					
	School_33	5.232	6.614	0.02	0.791	0.429
	School_34	-10.683	4.831	-0.069	-2.211	0.027
3	(Constant)	415.569	3.519		118.086	0
	School_1	-2.514	4.416	-0.023	-0.569	0.569
	School_2	-10.645	5.031	-0.068	-2.116	0.034
	School_3	5.314	4.124	0.052	1.288	0.198
	.					
	School_33	2.333	6.703	0.009	0.348	0.728
	School_34	-14.143	5.014	-0.092	-2.821	0.005
	TX	4.016	1.59	0.07	2.525	0.012

a. Dependent Variable: OAT_09

It is our contention that HLM has been frequently compared to MLR models that are often incorrectly written because they do not contain what we call "person vectors" or in the above case "school vectors." If one looks at the HLM Level 1 and Level 2 models presented below, it is apparent that they are much more similar to Model 2 above (see the output below that reflects this).

Multileveled Modeling

The Multileveled Models/Linear Mixed Models (LMM) reflect the same question that was tested by the regression models containing school vectors. In this case, the actual mixed model is created by substituting B_{0j} with B_0+u_{0j} and B_{1j} is replaced with y_{10}, but the error u_{1j} is not included since the differential treatment effects across schools are not being tested. This fourth set of models represents the LMM and the combined mixed model.

Model 4: LMM
Level 1 (Student): $OAT\_9 = B_{0j}+B_{1j}(TX_i)+r_{ij}$
Level 2(School) $B_{0j}=y_{00}+u_{0j}$
                              $B_{1j}=y_{10}+u_{1j}$

Mixed Model: $OAT\_9 = y_{00}+u_{0j}+y_{10}(TX)+r_{ij}$

Table 4 displays the estimates of the overall goodness of fit for this model looking at the -2 Restricted Log Likelihood statistic. Assessing fit and comparing different multilevel models becomes necessary in calculating the chi-square statistic.

χ^2= (-2 Restricted Log Likelihood$_{old}$) (-2 Restricted Log Likelihood$_{New}$)
df= Number of Parameters$_{old}$ – Number of Parameters$_{New}$

Table 4. Goodness of Fit Estimates

Information Criteria^a

-2 Restricted Log Likelihood	16944.349
Akaike's Information Criterion (AIC)	16950.349
Hurvich and Tsai's Criterion (AICC)	16950.362
Bozdogan's Criterion (CAIC)	16969.810
Schwarz's Bayesian Criterion (BIC)	16966.810

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: OAT_09.

The other goodness of fit estimates (e.g., AIC, AICC, CAIC, BIC) all adjust for different model complexities such as the number of parameters in the model and the sample size.

The Type III Test of Fixed Effects gives the overall fixed effect for TX and schools. In this case, only TX is of interest because schools are considered to be random. It should be noted at this point that Treatment is a fixed effect. It is also possible to have Treatment as a random effect depending on how it is operationally defined. In this case, the student either received treatment or did not. As can be seen, the p-value for the fixed effect of TX (p = 0.012) is the same as the p-value calculated when testing Model 2 against Model 3.

Table 5. Fixed Effects Coefficients

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	1785	313550.052	.000
TX	1	1785	6.376	.012
s_irn_2009	34	1785	9.241	.000

a. Dependent Variable: OAT_09.

Table 6 displays the coefficients of the fixed effects. We were not interested in the fixed effects of schools because not all of the schools were included in this illustration. As one can see from Table 6, the partial regression coefficient for TX is = -4.0159, which is again the same as that in the regression model with school vectors.

As noted earlier, the question of interest is "Does treatment account for a significant proportion of unique variance in predicting OAT independent of school differences?" Students who received the treatment scored, on average, 4.0159 points higher on the OAT than their comparison group. Even though this is a small difference (R^2_{change}=0.003), when either the LMM or regression is specified correctly, the power to detect these types of differences in nested design data becomes possible and equivalent. *Type VI Error*

A Type VI Error is a catchall concept that describes the inconstancy between the research question of interest and the statistical model (Newman et al., 2002; Tracz et al., 2005; Tracz et al., 2004). Too often these inconsistencies are overlooked leading researchers to make incorrect inferences. In addition to specifying the appropriate model, here are some other areas related to multilevel modeling that may be related to Type VI Errors.

Table 6. Fixed Effects Estimates

Estimates of Fixed Effects[b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	419.584607	3.850561	1785	108.967	.000	412.032526	427.136689
[TX=.00]	-4.015916	1.590371	1785	-2.525	.012	-7.135100	-.896732
[TX=1.00]	0	0	.	.	.	.	.
[s_irn_2009=257.00]	-2.514425	4.416500	1785	-.569	.569	-11.176479	6.147629
[s_irn_2009=352.00]	-10.644815	5.030888	1785	-2.116	.034	-20.511864	-.777766
[s_irn_2009=388.00]	5.314262	4.124444	1785	1.288	.198	-2.774985	13.403509
[s_irn_2009=22012.00]	-11.446891	5.787912	1785	-1.978	.048	-22.798689	-.095094
[s_irn_2009=23192.00]	2.332588	6.703082	1785	.348	.728	-10.814125	15.479301
[s_irn_2009=23283.00]	-14.143284	5.014410	1785	-2.821	.005	-23.978016	-4.308553

Centering

Centering is used most frequently when running mixed models or HLM. It is simply subtracting the mean from each score so that the mean of the distribution becomes zero. The choice to center is not a simple mathematical or statistical decision. It should be based upon the researcher's question of interest and/or theoretical position. There are three major decisions one has to make about centering: 1). Should one center? 2). If centering, should grand mean centering be used? 3). Should one use group mean centering? Grand mean centering is generally preferred over group mean centering (Burton, 1993; Hoffman & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995). Sarkisian (2007) takes the position that the original metric should never be used if the value of zero is not meaningful, and finds that there is a lack of precision in estimating the intercept in HLM when one does not center. According to Field (2009), centering is not an easy decision. It requires an understanding of the data and the analysis. Field also suggests centering may be a useful way to ameliorate the problem of multicollinearity between independent variables, especially when the independent variable does not have an interpretable zero value. Field points out that it is important to note that when using the group mean centering approach, the group mean should be considered a second level variable whenever group effects are not of interest. This situation is encountered frequently when an independent variable, such as time, is of interest.

If the researcher is interested in the relative position of the subject with regard to the treatment group mean, then group mean centering should be used. If, on the other hand, the researcher is interested in the absolute value of the independent variable (predictor variable), then grand mean centering should be used. When one does grand mean centering, the intercept becomes the adjusted grand mean. This adjustment obviously does not have any effect on the slopes. When group mean centering is used, the intercept is interpreted as the mean of each group. Group mean centering may change the meaning of the coefficient so that it becomes difficult to interpret because the mean values are subtracted from different sets of raw data. Some researchers will even center group mean binary variables; one needs to keep in mind that with group centered predictor variables, only person level effects are estimated. Choosing to center or not to center, and determining whether to use grand or group means, relates to Type VI Error because each decision will affect the statistical model that will differentially reflect the research question of interest.

Interaction

The classical definition of interaction is the differential effect across an area of interest (non-equal slopes) over and above the main effect (or controlling for the main effect). If looking at HLM to see if the second level accounts for a significant proportion of variance, one is looking at a differential effect across the area of interest, but it is not over and above the main effects. In other words, only the multiplicative slope differences are being tested, but not the slope differences independent of the main effects. Comparing the first level to second level HLM models is very similar to traditional interaction, but because it does not include the main effects, the results could be different.

Adequacy of Sample Size

Another important issue that is too often overlooked when using HLM is the adequacy of the sample size, especially as it relates to the higher-level variables in a model. Kreft (1996) found that to have sufficient power one needs at least 30 groups with 30 subjects per group, or 60 groups with 25 replicates per group, or 150 groups having 5 replicates per group. Kreft's simulation data suggests that the number of groups is more important than the number of observations for statistical power. This potential lack of power, from not having enough groups, has implications for detecting interaction between levels. Hox (1995) and Hox and Maas (2001) have similar findings related to adequacy of sample size. They found that N < 20 is insufficient at the higher levels, and if these higher-level variables are crucial to the structural model, then N should be > 100. These results are not consistent with the position taken by Raudenbush and Bryk (2002) who believe a Bayesian estimation approach allows for smaller N.

Research Design (Fixed Effects and Randomized Effects)

One of the major issues related to HLM and OLS/MLR is that the researcher has to determine if the design consists of fixed, mixed, or random effects. It is important to know the nature of these effects so that one can select the appropriate error terms. If the correct error term is not selected, the researcher cannot determine the correct error rates for the tests of significance.

Fixed effects occur when the variables of interest are assumed to not be randomly selected and no generalizations are going to be made beyond the variables being tested. For example, if there are three treatments that a researcher is interested in testing to see if they have a differential effect, then only those specific three treatments would be tested. This variable is fixed. In another example, imagine that a researcher is interested in the effects of a range of drug doses and randomly selects three doses to be representative of the entire range. Because the researcher wants to infer to the whole range of doses from which the samples are drawn, the three levels of the selected doses are considered to be random effects. In mixed effects, one must have at least two independent variables with at least one fixed and one random.

When building regression or hierarchical linear models, different error terms are used to test for statistical significance depending upon whether the variables are fixed or random. Being able to determine if the variables are fixed or random is important in testing models because they have different assumptions. In our opinion, one important issue here is conceptual; not statistical. As an example, assume a researcher is interested in having drug dosage as a continuous variable, and interested in generalizing to the whole range of dosages (i.e., think of this as a variable on the X axis of a graph). In this scenario, it would appear that dosage is a continuous variable. If, however, the researcher were to take drug dosage and was only interested in generalizing to three categories (i.e., small, medium, or large doses) based upon some predetermined decision rule, then the variable dosage would have changed from a continuous (random) variable to a categorical (fixed) variable. Obviously, with the fixed variable the researcher is specifically addressing whether there is a difference between the small, medium, and large dosage levels and not attempting to generalize to the range of doses. The robustness of violations to the underlying assumptions of the fixed and randomized models, as they relate to the accuracy of the tests of significance, is considerable. The fixed model, especially when the design is balanced, is more robust than the randomized model.

Conclusion

The authors contend that there is nothing more important than understanding the relationship between the question of interest, the data, and the analysis. There is no computer program that is capable of doing the researcher's thinking for him or her. Therefore, it could be very misleading to use default options on computer programs or to use very sophisticated computer programs that have algorithms that are virtually black boxes and are too often not understood by the researcher. HLM is being widely used to analyze data, but we are concerned about how well these models are understood and appropriately interpreted. This concern is especially true when considering that stability is heavily based upon the number of replicates at the higher-order levels.

HLM modeling for the above example frequently takes into consideration the differential effect of treatment for different schools (a "type of interaction"). However, this question was not the one that was posed in the above OLS models. One could run models to reflect interaction, but that would go beyond the purpose of this paper. We have demonstrated that when models were written to reflect question of

interest ("Was there an overall different effect due to treatment, independent of schools?"), the results of OLS and HLM were virtually identical.

To complicate matters, there is an option in HLM to run treatment as a fixed or random effect. If treatment is thought of as a random effect, (this choice would not make sense in the above example), and if HLM were run using treatment as a random effect, the results would be different from those obtained using OLS modeling. Additionally, various "experts" in HLM suggest different approaches for writing the HLM models in their reference books. These diverse approaches may produce dissimilar results (see Bickel, 2007; Field, 2009; Raudenbush & Bryk, 2002). The choice to use random or fixed effects, the type of centering selected, etc. needs to be contingent upon the researcher's understanding of the data and the purpose of the research. While there are advantages to using HLM, researchers must be aware that it is not always the most efficacious procedure. One must select the method that will most accurately reflect the research question in the simplest way. This selection is analogous to the concept of parsimony.

References

Bickel, R. (2007). *Multilevel analysis for applied research: It's on regression.* New York: The Guilford Press.

Burton, B. (1993). *Some observations on the effect of centering on the results obtained from hierarchical linear modeling*. Washington, D.C.: National Center for Education Statistics, U. S. Department of Education.

Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage Publication.

Hoffman, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research organizations. *Journal of Management, 24*, 623-641.

Hox, J. J. (1995). *Applied multi-level analysis*: *A basic, non-technical introductory text* (2nd ed.). Amsterdam, Netherlands: TT-Publikaties.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8,* 157-174.

Kreft, I. G. G. (1996). *Are multi-level techniques necessary? An overview, including simulation studies*. Retrieved November 22, 1999, from http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research 30*, 1-21.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model*. Carbondale, IL: Southern Illinois University Press.

Morris, C. (1995). Hierarchical models for educational data- an overview. *Journal of Educational and Behavior Statistics, 20*, 190-200.

Mundfrom, D. J., & Schults, M. R. (2002). A monte carlo simulation comparing parameter estimates from multiple linear regression and hierarchical linear modeling. *Multiple Regression Viewpoints, 28*, 18-21.

Newman, I., Fraas, J., Newman, C., & Brown, R. (Fall, 2002). Research practices that produce Type VI Errors. *Journal of Research in Education, 12*, 138-145.

Pedazur, E. J. (1982). *Multiple regression and behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart & Winston.

Raudenbush, S. W. (2009) Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis*. (pp. 295- 315). New York: Russell Sage Foundation.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Sarkisian, N. (2007). *HLM model building strategies: Class notes*. Retrieved April 2, 2010, from www.sarkisian.net/sc705/september6.pdf

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Tracz, S., Nelson, L., Newman, I., & Beltran, A. (2005). The misuse of ANCOVA: The academic and political implications of Type VI errors in studies of achievement and socioeconomic status. *Multiple Linear Regression Viewpoints, 31*, 16-21.

Tracz, S., Newman, I., Nelson, L., & Dellran, A. (2004, October). *How ANCOVA can be misused in studies of achievement and socioeconomic status: The academic and political implications of Type VI*

errors. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

Williams, J. D. (1987). The use of nonsense coding with ANOVA situations. *Multiple Linear Regression Viewpoints 15*, 29-39.

Send correspondence to: David Newman
                            Cleveland State University
                            Email: d.o.newman@csuohio.edu

Regression Commonality Analysis:
Demonstration of an SPSS Solution

Kim Nimon
University of North Texas

Multiple regression is a widely used technique to study complex interrelationships in social science research. In the face of multicollinearity, researchers encounter challenges when interpreting multiple linear regression results. Although beta weights and structure coefficients provide insight into the synthetic variable ($\hat{y}$) produced, they fall short when researchers want to fully report regression effects. Regression commonality analysis provides a level of interpretation of regression effects that cannot be revealed when only examining beta weights and structure coefficients. Importantly, commonality analysis provides a full accounting of regression effects which identifies the loci and effects of suppression and multicollinearity. Conducting regression commonality analysis without the aid of software is laborious and may be untenable, depending on the number of predictor variables. A software solution is presented for the multiple regression case.

Multiple regression is a powerful technique for identifying underlying complex correlations among data in social science research. However, when multiple predictor variables are introduced into a regression model, the interpretation of the results is not always easy. It is often the case that predictor variables are multicollinear (i.e., have nonzero correlations with each other), thus increasing the complexity of the covariance structure and the interpretation of the results. Depending on the extent, researchers may see multicollinearity as a problem they need to "fix" before conducting multiple regression. However, Henson (2002) noted that multicollinearity is not a problem in multiple regression as long as researchers use appropriate methods to investigate their results. As noted by Nimon and Roberts (2010), some of these methods include, but are not limited to, the investigation of beta weights, structure coefficients, and commonality coefficients.

While analyzing beta weights is somewhat a de facto process when conducting multiple regression, Courville and Thompson (2001) demonstrated the ill effects of relying solely on beta weights to evaluate predictor importance. As they noted, beta weights are affected by the level of multicollinearity with other predictors and by themselves are not good indicators of predictor usefulness. Furthermore, researchers may miss the presence of suppressor variables if they neglect to analyze beta weights in concert with structure coefficients.

However, analyzing beta weights and structure coefficients may not provide sufficient information to fully interpret a regression effect since these coefficients do little to identify the loci and magnitude of multicollinearity and suppression. For example, a predictor (i.e., X1) with a high beta weight and a low structure coefficient indicates that X1 serves as a suppressor variable in the regression effect. However, without further analysis of the regression effect, it is not apparent which variable(s) is the target of suppression nor is the magnitude of the suppression effect clear.

By computing commonality coefficients, a predictor's contribution to a regression effect can be related to the other predictor variables in the model. Such information can be useful for uncovering complex relationships and for informing theory. For example, Siebold and McPhee (1979) discovered that cognition served as a suppressor variable in predicting the relationship between social affect and intent to get a pap test among minority women. As they noted, "a message aimed at social factors is likely to have maximum impact on intentions if it can be `purified' of cognitive relevance" (p. 365). Without analyzing commonality coefficients, this relationship would not have been revealed. An additional benefit of computing commonality coefficients is that they can be used to compute squared structure coefficients where, according to Nimon, Lewis, Kane, and Haynes (2008) "The benefit of employing commonality analysis in conjunction with the analysis of squared structure coefficients is that the researcher can determine how much variance each variable uniquely contributes and how much each shares, if any, with every other variable in the regression" (pp. 460-461).

Commonality Analysis

Commonality coefficients are produced through the process of commonality analysis[1]. Promoted in the 1960s as a method of partitioning variance (R^2) (Mayeske et. al, 1969; Mood, 1969, 1971; Newton & Spurrell, 1967), commonality analysis provides a method to determine the variance accounted for by respective predictor variable sets (Onwuegbuzie & Daniel, 2003; Rowell, 1996) and helps researchers understand the contributions predictor variables make in a given regression model (Zientek & Thompson, 2009). Commonality analysis has been applied across disciplines in social science research, including education (e.g., Zientek & Thompson, 2006), counseling (e.g., Gill, Barrio Minton, & Myers, 2010), human resource development (c.f., Nimon, Gavrilova, & Roberts, 2010), behavioral science (e.g., Sorice & Conner, 2010), and information science (e.g., Nimon & Gavrilova, 2010). Across these disciplines and others, data resulting from commonality analysis provide rich interpretation of the regression effect that advances theory and the application of research findings. As Siebold and McPhee (1979) stated:

> Advancement of theory and the useful application of research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to which those independent variables, singly and in all combinations, share variance with the dependent variable. Only then can we fully know the relative importance of independent variables with regard to the dependent variable in question. (p. 355)

Commonality analysis partitions a regression effect into constituent, non-overlapping parts (Thompson, 2006). The partitioning process produces unique and common effects. Unique effects identify how much variance is unique to an observed variable, and common effects identify how much variance is common to two or more variables. Unique effects are analogous to squared semi-partial correlations between a predictor and dependent variable. Unique effects are also equivalent to the increase in R^2 that occurs when a variable is included by itself in the last block of a hierarchical regression. Although hierarchical regression can be used to compute specific unique effects, it does not produce sufficient $R^2 s$ to compute common effects.

Consider the hypothetical situation discussed by Hedges and Olkin (1981) in which two variables, X1 and X2, are used to predict a variable X0. As depicted in Figure 1, the variance in X0 that is explained by X1 and X2 ($R^2_{0.12}$) can be partitioned into three components:

$$\gamma_1 = \text{unique effect of X1 to } R^2_{0.12}$$
$$\gamma_2 = \text{unique effect of X2 to } R^2_{0.12}$$
$$\gamma_{12} = \text{common effect of X1 and X2 to } R^2_{0.12}.$$

and computed as follows:

$$\gamma_1 = R^2_{0.12} - R^2_{0.2}$$
$$\gamma_2 = R^2_{0.12} - R^2_{0.1}$$
$$\gamma_{12} = R^2_{0.1} + R^2_{0.2} - R^2_{0.12.}$$

Had a hierarchical regression been conducted where X1 were entered in the first block followed by X2 in the second block, the resulting analyses would produce $R^2_{0.1}$ and $R^2_{0.12}$. While these $R^2 s$ are sufficient to compute γ_2, they are insufficient to compute γ_1 and γ_{12}. Although hierarchical regression is valuable in identifying the delta in R^2 between variables sets, it does not provide sufficient $R^2 s$ to fully partition a regression effect into constituent, non-overlapping parts.

Considering the prior example with only two predictor variables, the commonality analysis procedure is fairly straightforward. However, as the number of predictor variables increases, the process becomes more complicated. This is due, in large, to the fact that the number of commonality coefficients is exponentially related to the number of predictor variables. For example, the number of commonality coefficients for three, four, or five predictor variables is 7, 15, and 31, respectively. In addition, the formulas to compute commonality coefficients differ according to the number of predictors in the model. While complete sets of commonality formulas have been published supporting up to four predictor variables (e.g., Frederick, 1999), formulas to support models with more than four predictor variables must be derived. Mood's (1969) procedure allows researchers to calculate commonality coefficient formulas for any number of predictor variables. However, the procedure can be somewhat difficult to implement across a large number of coefficients. Nimon et al. (2008) summarized the procedure as follows:

> In Mood's (1969) procedure, (*1-x*) was used to represent variables in the common variance subset and (*x*) was used to represent variables not in the common variance subset. By negating

the product of the variables in the subset and the variables not in the subset, deleting the -1 resulting from the expansion of the product, and replacing x with R^2, Mood noted that the formula for computing any commonality coefficient can be derived. (p. 459)
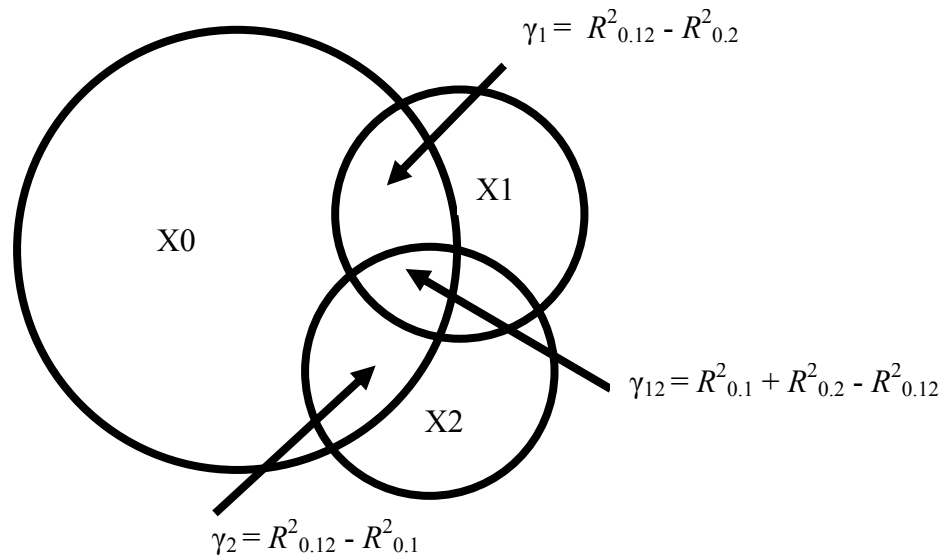

$$\gamma_1 = R^2_{0.12} - R^2_{0.2}$$

$$\gamma_{12} = R^2_{0.1} + R^2_{0.2} - R^2_{0.12}$$

$$\gamma_2 = R^2_{0.12} - R^2_{0.1}$$

Figure 1. Venn diagram of the unique (γ_1, γ_2) and common (γ_{12}) effects of $R^2_{0.12}$.

For example, to compute the variance common to two independent variables (X1, X2) out of two (i.e., γ_{12}), Mood's procedure results in the following:

$$-(1 - X1)(1 - X2) = -(1 - X1 - X2 + X12) = -X12 + X1 + X2 = -R^2_{12} + R^2_1 + R^2_2$$

Traditionally, the process of computing commonality coefficients has been a manual process[2] involving three steps: (1) developing the appropriate commonality coefficients formulae, (2) executing an all possible subsets (APS) regression, and (3) populating the commonality coefficient formulae with the appropriate R^2 values from the APS regression. Without the aid of software, this process can be laborious and even untenable, depending on the number of predictor variables.

Through the work of Nimon et al. (2008), users of R have the opportunity to report commonality coefficients for any number of predictor variables by simply invoking a function from the yhat package (Nimon & Roberts, 2009). R is a free statistical programming language and environment for the Unix, Windows, and Mac families of operating systems (R Development Core Team, 2010). Despite R's growing popularity, researchers may prefer to conduct multiple regression using more traditional statistical software packages such as Statistical Package for the Social Sciences (SPSS), Statistical Analysis System (SAS), and SYSTAT. To provide commonality analysis functionality to SPSS users, the work of Nimon et al. (2008) was migrated to an SPSS script that conducts commonality analysis for any number of predictor variables.
The SPSS script file can be downloaded from http://profnimon.com/commonality.sbs.

SPSS Commonality Demonstration
To demonstrate the functionality of the SPSS script, the heuristic dataset from Thompson (2006, p. 221) is employed in which three predictors (X4, X5, and X7) are regressed on a dependent variable Y. As depicted in Figures 2 to 5, the script file prompts the user for the: (1) SPSS data file, (2) output filename prefix, (3) dependent variable, and (4) independent variables. Due to limitations in the SPSS MATRIX command, all user input (data file name, output filename prefix, and variable names) must be eight characters or less.

Figure 2. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 1.

Figure 3. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 2.

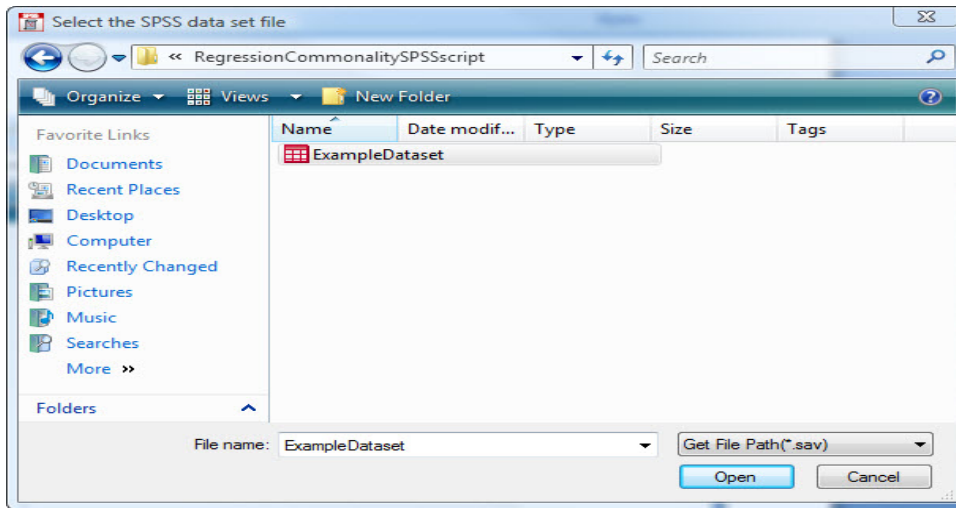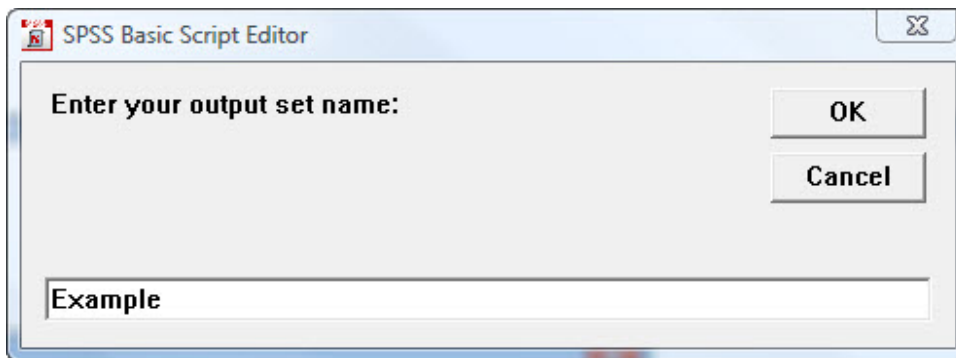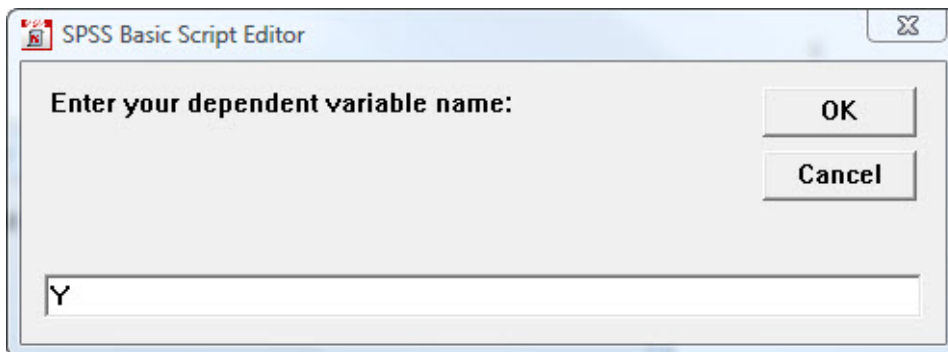**Figure 4**. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 3.
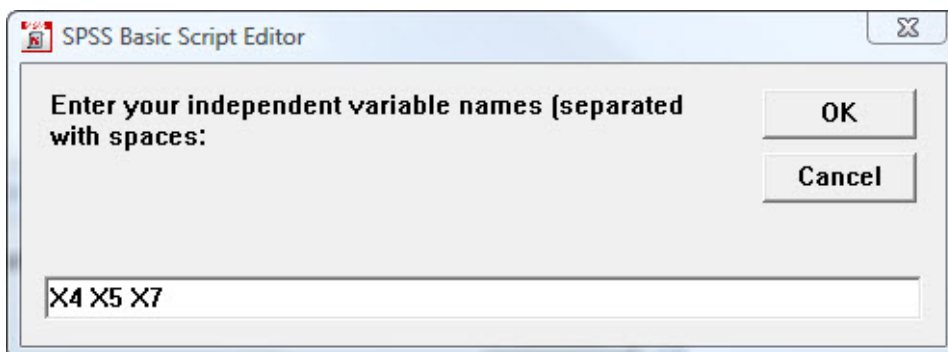

Figure 5. Screen Snapshot of Regression Commonality SPSS Script User Input – Step 4.

Using the information supplied, the script generates two SPSS data files - CommonalityMatrix.sav and CCByVariable.sav where both file names are prepended with the output file name prefix and are placed in the same directory as the input SPSS data file. CommonalityMatrix.sav contains the unique and common commonality coefficients as well as the percent of variance in the regression effect that each coefficient contributes. The individual entries in the table can be used to determine how much variance is explained by each effect as well as which coefficients contribute most to the regression effect. CCByVariable.sav provides another view of the commonality effects. The unique effect for each of the predictors is tabularized, as well as the total of all common effects for which the predictor is involved. The last column is the sum of the unique and common effect and is equivalent to the squared correlation between the predictor and dependent variable. As such, dividing the variance sum by the regression effect yields the percent variance explained by each variable, equivalent to a squared structure coefficient (r_s^2).

Tables 1 and 2, respectively, contain the contents of ExampleCommonalityMatrix.sav and ExampleCCByVariable.sav for the example regression model. Table 3 presents an example of how the commonality effects by variable can be displayed alongside traditional multiple regression output to add another layer of consideration when evaluating the importance of predictors.

Table 1. ExampleCommonalityMatrix

Variables	Coefficient	% Total
Unique to X4	0.251	45.925
Unique to X5	0.308	56.274
Unique to X7	0.056	10.199
Common to X4 X5	-0.042	-7.738
Common to X4 X7	0.156	28.465
Common to X5 X7	-0.049	-8.947
Common to X4 X5 X7	-0.132	-24.177
Total	0.547	100.000

Table 2. ExampleCCbyVariable

Variable	Unique	Common	Total
X4	.251	-.019	.232
X5	.308	-.223	.084
X7	.056	-.025	.030

Table 3. Regression Results for Heuristic Dataset

Predictor (x)	R	R^2	R^2_{adj}	β	p	Unique	Common	Total	% of R^2 (r_s^2)
	.742	.547	.462						
X4				.584	.009	.251	-.019	.232	42.225%
X5				.717	.005	.308	-.223	.084	15.321%
X7				.329	.180	.056	-.025	.030	5.507%

Note:. *Unique = x*'s unique effect. *Common = Σ x*'s common effects. *Total = Unique + Common.* % of R^2 = *Total/R^2*

As seen in Table 1, regressing X4, X5, and X7 on Y resulted in negative commonality coefficients. Negative commonality coefficients occur in the presence of suppression or when predictors affect each other in the opposite direction (Pedhazur, 1997). Negative commonality coefficients indicate that one variable actually confounds the predictive power of another (Beaton, 1973). While Frederick (1999) indicated that negative commonalities should be interpreted as zero, others have disagreed (e.g., Beaton; Capraro & Capraro, 2001; Pedhazur, 1997). Their magnitude indicate the power (i.e., variance explained) that is achieved by including the confounding variable (Capraro & Capraro, 2001). Consider the example from Capraro and Capraro that demonstrates this phenomenon:

An Olympic track athlete must be fast and strong, therefore, a strong-fast athlete would be correlated with success at running track. However, one would believe the two variables (fast and strong) would be moderately negatively correlated, that is as much strength and mass increases, speed would decrease. The negative commonality between speed and strength would indicate a confounded variable. In this case, by knowing both the speed and strength one would expect to make better predictions of successful track running. Imagine just knowing the speed

or strength of the athlete. A fast athlete may perform well in a short sprint but be severely impaired in a distance event. Conversely, a strong athlete may excel in endurance and persevere for distance, but lack the speed to win. The negative commonality in this case indicates that the power of both variables is greater when the other variable is also used. (pp. 6-7)

The commonality data in Table 1 indicate that the regression effect was confounded by all of the predictor variable combinations involving X5. Including X5 in the regression model increased X4's unique contribution by .042, X7's unique contribution by .049, and X4 and X5's common contribution by .132. Given its correlation with Y ($r = .290$), it appears that X5 served as a partial suppressor variable. Note that the magnitude of the change in R^2 when X5 is dropped from the equation (.308) far surpasses its relationship with the dependent variable (.084). In total, suppression accounted for 40.87% (7.74% + 8.95% + 24.18%) of the regression effect.

The predictor X5 was also the major contributor to the regression effect. Not only did it contribute 40.87% of the regression through its role in suppressing variance in X4 and X7 that was irrelevant to predicting Y, it uniquely contributed 15.40% of the regression effect (56.27% - 40.87%). The other major contributor to the regression effect was the unique variance associated with the predictor X4. Excluding its relationship with X7, X4 uniquely contributed 38.17% of the regression effect.

Table 3 presents one format for reporting commonality data alongside traditional regression results. Such a table allows researchers to simultaneously consider beta weights, structure coefficients, unique effects, and common effects when interpreting regression effects and predictor importance. For instance, since the sum of the squared structure coefficients (42.23% + 15.32% + 5.51%) in the example regression model is less than 100%, it is clear that the regression effect is confounded by suppression. Examining the unique and common effect identifies X5 as a suppressor variable and helps resolve the discrepancy between its large beta weight and small squared structure coefficient. One might also observe the role that X7 plays in the regression effect. Not only does it have a small beta weight and small structure coefficient, its unique effect indicates that it could be excluded from the regression model with only a small reduction in R^2 (.547 - .056 = .491).

Conclusion

From a didactic perspective, commonality analysis clarifies the roles that multicollinearity and suppression play in the relationship between standardized function and squared structure coefficients. In addition, it can be observed that commonality analysis subsumes the role of computing squared structure coefficients because the portion of the regression effect explained by each variable generated from the regression commonality analysis is identical to the squared structure coefficient generated from multiple linear regression. From a theoretical perspective, regression commonality analysis can provide important insights into variable relationships that may not be revealed by only examining beta weights and structure coefficients. Researchers and practitioners should consider employing one of the commonality analysis software solutions identified in this paper and others that may evolve (e.g., SAS, SYSTAT) when interpreting multiple linear regression results.

Notes

1. A complete discussion of regression commonality analysis is beyond the scope of this article; readers are referred to accessible treatments of the topic by Amado (2003), Mood (1971), Pedhazur (1997), Rowell (1996), Seibold and McPhee (1979), Thompson (2006), and Zientek and Thompson (2006). Commonality analysis as defined in this paper stems from the seminal works of Beaton (1973); Creager and Valentine (1962); Kempthorne (1957); Mayeske et al. (1969); Mood (1969); and Newton and Spurrell (1967). As defined by these authors, it is a R^2 type variance partitioning analysis. The software embodiment of regression commonality analysis described in this paper is distinct and different than the commonality analysis software distributed by International Business Machines Corporation (2006) that identifies pairs of columns that have a significant number of common domain values.

2. SAS software (SAS Institute Inc., 2008) and SYSTAT (2004) support APS regression, which is one of the steps in conducting regression commonality analysis. Also, a FORTRAN IV computer program to accomplish commonality analysis was introduced by Morris (1976). However, this program is now obsolete as it requires input job control cards.

References

Amado, A. J. (2003). Partitioning predicted variance into constituent parts: A primer on regression commonality analysis. *Research in the Schools, 10*(1), 91-97.

Beaton, A. E. (1973, March). *Commonality.* (ERIC Document Reproduction Service No. ED111829)

Capraro, R. M., & Capraro, M. M. (2001). Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. *Multiple Linear Regression Viewpoints*, *27*(2), 16-23.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: ß is not enough. *Educational and Psychological Measurement, 61*, 229-248. doi: 10.1177/0013164401612006.

Creager, J. A., & Valentine, L. D. Jr. (1962). Regression analysis of linear composite variance. *Psychometrika*, *27*, 31-37.

Frederick, B. N. (1999). Partitioning variance in the multivariate case: A step-by-step guide to canonical commonality analysis. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 305-318). Stamford, CT: JAI Press.

Gill, C. S., Barrio Minton, C. A., & Myers, J. E. (2010). Spirituality and religiosity: Factors affecting wellness among low-income, rural women. *Journal of Counseling & Development, 77*, 293-303.

Hedges, L. V., & Olkin, I. (1981). The asymptotic distribution of commonality components. *Psychometrika*, *46*, 331-336.

Henson, R. K. (2002, April). *The logic and interpretation of structure coefficients in multivariate general linear model analyses*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED467381)

International Business Machines Corporation (2006). *IBM InfoSphere information analyzer: Analysis methodology and best practices guide, Version 8 Release 1.2*. Retrieved 7 July, 2010, from http://publibfp.boulder.ibm.com/epubs/pdf/c1927501.pdf

Kempthorne, O. (1957). *An introduction to genetic statistics*. New York: John Wiley.

Mayeske, G. W., Cohen, W. M., Wisler, C. E., Okada, T., Beaton, A. E., Proshek, J. M., Weinfeld, F. D., & Tabler, K. A. (1969). *A study of our nation's schools*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Mood, A. M. (1969). Macro-analysis of the American educational system. *Operations Research, 17*, 770-784.

Mood, A. M. (1971). Partitioning variance in multiple regression analyses as a tool for developing learning models. *American Educational Research Journal*, *8*, 191-202.

Morris, J. D. (1976). A computer program to accomplish commonality analysis. *Educational and Psychological Measurement, 36*, 721-723.

Newton, R. G., & Spurrell, D. J. (1967). A development of multiple regression for the analysis of routine data. *Applied Statistics, 16*, 51-64.

Nimon, K., & Gavrilova, M. (2010, February). *Commonality analysis: Demonstration of an SPSS solution for regression analysis*. Poster presented at the 2010 iConference, University of Illinois at Urbana-Champaign.

Nimon, K., Gavrilova, M., & Roberts, J. K. (2010). Regression results in human resource development research: Are we reporting enough? In C. Graham & K. Dirani (Eds.), *Proceedings of the Human Resource Development 2010 International Conference* (pp. 803-812), Knoxville, TN: AHRD.

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*, 457-466. doi: 10.3758/BRM.40.2.457

Nimon, K., & Roberts, J. K. (2009). *yhat: Interpreting Regression Effects*. R package version 1.0-2.

Nimon, K., & Roberts, J. K. (2010). *Interpreting linear models: The importance of beta weights, structure coefficients, commonality coefficients, and adjusted effect sizes*. Manuscript submitted for publication.

Onwuegbuzie, A. J., & Daniel, L. G. (2003, February 19). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education* [On-line], 6(2). Available: http://cie.ed.asu.edu/volume6/number2/

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Ft. Worth, TX: Harcourt Brace.

R Development Core Team. (2010). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Retrieved 7 July, 2010, from http://www.R-project.org (ISBN 3-900051-07-0)

Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct regression commonality analysis. *Advances in Social Science Methodology, 4*, 33-43.

SAS Institute Inc. (2008). *SAS/STAT® user's guide, Version 9.2*. Cary, NC: Author.

Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analysis. *Human Communication Research, 5*(4), 355-363.

Sorice, M. G., & Conner, J. R. (2010). Predicting private landowner intentions to enroll in an incentive program to protect endangered species. *Human Dimensions of Wildlife*, *15*(2), 77-89.

Systat Software, Inc. (2004). *SYSTAT 11 language reference user guide.* Chicago, IL: Author.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: The Guilford Press.

Zientek, L. R., & Thomspon, B. (2006). Commonality analysis: Partitioning variance to facilitate better understanding of data. *Journal of Early Intervention, 28*(4), 299–307.

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher, 38*, 343-352.

Send correspondence to: Kim Nimon
                            University of North Texas
                            Email: kim.nimon@unt.edu

A Confirmatory Factor Analysis of
the Attitudes Toward Research Scale

David A. Walker
Northern Illinois University

A confirmatory factor analysis (CFA) was used to examine if the developed five-factor structure of the Attitudes Toward Research scale was a proper model fit or if a different model more appropriately fit the sample data. The CFA conducted suggested that the three-factor model supported the sample data better than the one- or the five-factor structures, and numerous model fit indices comparison results indicated that the three-factor model was the more preferred model. The three-factor model was comprised of the latent factors: 1) research use, 2) negative attributes of research, and 3) positive attributes of research.

A problem encountered by faculty, often derived from instructional experience or tacit professional expertise, with some students that have taken a course or courses in research methods is that years, or even a few semesters, after having been enrolled, they sometimes do not practice solid research design when engaged in the various phases of the master's thesis and/or doctoral dissertation process. Beyond a lack of experience with research or time passage between courses and thesis or dissertation work, two leading indicators related to this limitation of research preparedness may be due to anxiety (Onwuegbuzie, 2003) affiliated with research methods-like courses and negative attitudes toward these types of courses (Onwuegbuzie, 1997; Wise, 1985). Further, this latent problem may be discernible earlier in research methods course(s) from enrolled students as constructs of their concerns pertaining to research's lack of usefulness (i.e., misunderstandings of interpreting statistical findings or lack of job support for research use); a want of understanding, or only rudimentary, introductory awareness, concerning the importance of research in the graduate education process (i.e., seen as program-imposed, mandatory classes); deficient comprehension of the relevance of research in professional endeavors, or self-efficacy issues concerning ability and motivation to learn and perform research-related procedures (Henson, Hull, & Williams, 2010; Onwuegbuzie, DaRos, & Ryan, 1997; Pajares & Schunk, 2001; Ranis, 2003; Ravid & Leon, 1995).

A review of the scholarly literature indicates a dearth in the knowledge base concerning research related to education students' attitudes toward research. To be sure, the literature has a wealth of research pertaining to students' perspectives on statistics courses such as the courses' level of difficulty, issues of student anxiety, and the validation of scales to measure said perceptions (cf. Cashin & Elmore, 1997; Kennedy & McCallister, 2001; Mills, 2004; Onwuegbuzie, 2004; Schau, Stevens, Dauphinee, & Del Vecchio, 1995). Further, there are studies to support attitudes toward research from professional perspectives such as nurses (Björkström & Hamrin, 2001) or administrators and faculty (Combs, Bustamante, & Wilson, 2007; Tang & Chamberlain, 1997), but very little data in the area of education students particularly in research methods courses (cf. Murtonen, 2005; Papanastasiou, 2005; Ranis, 2003; Richardson & Onwuegbuzie, 2002).

Development and Dimensionality of the ATR

As a means to develop these latent constructs, one method may be derived via an instrument that measures education students' attitudes toward research, which would include factors such as negativity or research usefulness. However, in the scholarly literature, there are very few of these instruments in circulation pertaining to this particular issue. One of the more readily available and employed protocols for this purpose, Papanastasiou's (2005) "Attitudes Toward Research" (ATR) scale, has not endured a confirmatory examination of its five-factor structure to determine if it fits said structure or if different structure would better represent the ATR scale.

Originally, the ATR scale consisted of 56 questions based on a 1 (strongly disagree) to 7 (strongly agree) Likert-type scale. A principal components factor analysis with varimax rotation was conducted on all 56 indicators. Eleven factors were retained that accounted for 66.40% of the variance. Further analysis reduced the number of items on the ATR scale from 56 to 32, where items that "were not significantly related to the total score, or whose coefficient [via a score reliability analysis] was less than .50 were removed" (Papanastasiou, 2005, p. 18). Also, non-qualifying items that had factor loadings < .50 were deemed "insignificant" and dropped from the scale. A second factor analysis was conducted on the remaining 32 items of the scale. A five-factor solution was retained that accounted for 66.25% of the

variance. Thus, the final structure of the ATR scale consisted of 32 indicators that had five identified constructs: 1) usefulness of research in a career; 2) research anxiety; 3) positive attitudes toward research; 4) research relevance to life; and 5) research difficulty. The score reliability estimates derived via Cronbach's alpha (α) from the ATR's five subscales were usefulness of research in a career (9 items, α = .92); research anxiety (8 items, α = .92); positive attitudes toward research (8 items, α = .93); research relevance to life (4 items, α = .77); and research difficulty (3 items, α = .71) (Papanastasiou, 2005).

Methods

Purpose

The purpose of this study was to use a confirmatory factor analysis (CFA) and various indicators of fit to examine how closely Papanastasiou's (2005) five-factor model represented the sample data and also how multiple models' fit indices compared to one another. Based on a sample of students from the U.S., this study examined if the reported five-factor structure of the ATR was a proper model fit or if a different model, in terms of interpretable factor structures and data fit, more appropriately fit the sample data. It was the proposition of the current research that a model consisting of an oblique three-factor model, where inter-related factors to the concept of research would be comprised of: 1) research use, 2) negative attributes of research, and 3) positive attributes of research.

Model Examination

A one-factor model was created as a representation of the ATR scale as a uni-dimensional structure, where attitudes toward research were nested within one factor: attitudes. A five-factor model, as advanced in the literature by Papanastasiou (2005), was constructed that represented the 32-item ATR scale as a multi-dimensional structure consisting of the following independent yet correlated factors: usefulness of research in a career, research anxiety, positive attitudes toward research, research relevance to life, and research difficulty.

Finally, a three-factor model containing a total of 18 items and accounting for 56% of the variance was created from Papanastasiou's 32-item five-factor model (see Table 1). Items that were found to be highly correlated or have large standard errors were deleted because they did not appear to add to the breadth of coverage to construct domains. For the three-factor model, the ATR scale was conceptualized as a multi-dimensional structure comprised of three independent yet correlated components: research use, negative attributes of research, and positive attributes of research.

Procedure and Participants

Two hundred four students in a graduate education research methods course at a large, state university located in the Midwest participated in the study. The ATR scale was completed by students during the 14th or 15th week of a 16-week course in 8 sections of the course over one academic year. Participants responded to the ATR's items based on a Likert-type scale ranging from 1 = Strongly Disagree to 7 = Strongly Agree. Only 2 out of the 204 completed questionnaires had missing data, where a respondent left 1 question blank and another participant left 2 different questions, in comparison to the former respondent, without responses. Because of minimal missing data, and the apparent lack of a pattern in the few missing data observed, the mean was imputed for these two participants' three missing data instances (cf. Baker & Siryk, 1999).

The sample, which was representative of the population of interest, was composed of 77.50% females and had an average age near 32 years (M = 31.95, SD = 8.46). The ethnic break-down of the sample was 83.80% Caucasian, 10.80 Latino/a, 3.40% African American, 1.50% Asian, and .50% Native American, which was very representative of this university's College of Education population of interest. Nearly all of the participants were seeking a master's degree (97.50%), with an average of 20 credits completed toward the degree. There were 17 majors represented in the sample, with the top three areas drawing students from: educational administration (16.70%), literacy (14.20%), and counseling (12.30%).

Table 1. Items on the ATR scale

Item
1. Research makes me anxious
2. Research should be taught to all students
3. I enjoy research
4. Research is interesting
5. I like research
6. I feel insecure concerning the analysis of research data
7. Research scares me
8. Research is useful for my career
9. I find it difficult to understand the concepts of research
10. I make many mistakes in research
11. I have trouble with arithmetic
12. I love research
13. I am interested in research
14. Research is connected to my field of study
15. Most students benefit from research
16. Research is stressful
17. Research is very valuable
18. Research makes me nervous
19. I use research in my daily life
20. The skills I have acquired in research will be helpful to me in the future
21. Research is useful to every professional
22. Knowledge from research is as useful as writing
23. Research is irrelevant to my life
24. Research should be indispensable in my professional training
25. Research is complicated
26. Research thinking does not apply to my personal life
27. I will employ research approaches in my profession
28. Research is difficult
29. I am inclined to study the details of research
30. Research acquired knowledge is as useful as arithmetic
31. Research-oriented thinking plays an important role in everyday life
32. Research is a complex subject

Note: All 32 items comprised the five-factor model. Bold and gray signify the 18 items used in the three-factor model

Data Screening

Univariate outliers were checked via boxplots and histograms, which indicated that the data were normal in this regard. Multivariate outliers were checked using the Mahalanobis distance statistic. The largest Mahalanobis d-squared value, or the observation the furthest distance from the centroid, was 46.70 with a probability value > .05 or a non-small probability of an unusual observation (i.e., outlier). Further, data were screened for instances of multicollinearity via analysis of tolerance (TOL) and variance inflation factor (VIF). Multicollinearity was not present as all TOL indices were > .10 and all VIF measures were < 3, which met noted cut-off points for these measures of > .10 and < 10, respectively (Belsley, Kuh, & Welsch, 1980; Hair, Anderson, Tatham, & Black, 1995).

Table 2 shows each item's mean, standard deviation, skewness, and kurtosis. In terms of standard deviation, there was a range from .98 to 1.81. Skewness ($\leq |1.21|$) and kurtosis ($\leq |1.69|$) results showed that none of the items were > the recommended cut-off points of $|3.00|$ and $|8.00|$, respectively, and there was no univariate non-normality present (Kline, 1998). Mardia's kurtosis value was used to check for multivariate normality. Mardia values as small as not > 3 and as large as not > 30 have been noted as a sign of multivariate kurtosis (Bentler & Wu, 1993; Newsom, 2005). The current study's Mardia value was 43, which was an indication that possibly there was multivariate kurtosis affiliated with the data. To

extend beyond the Mardia test for instances of multivariate non-normality, a Bollen-Stein (1992) bootstrap technique of 2,000 iterations (cf. Nevitt & Hancock, 2001) was conducted in AMOS 7.0 (Analysis of Moment Structures) (Arbuckle, 2006) to derive a p-value that was comparable to the p-value affiliated with the Satorra-Bentler (1994) chi-square adjustment used for model fit with data that exhibit non-normal tendencies. Thus, if the p-value from the Bollen-Stein bootstrap was > .05, we would fail to reject the model and determine that it indeed fit the data. Results indicated that the Bollen-Stein bootstrapped p-value was > .05 and the model appeared to fit the data. That is, the data comprising the models under study were robust in terms of a potential threat from multivariate non-normality. Thus, given the results from the various tests conducted pertaining to the normal distribution of the model's data, maximum likelihood estimation (MLE) was used.

Table 2. Three-factor model descriptive statistics

Item	M	SD	Skewness	Kurtosis
1. Research makes me anxious	4.40	1.81	-0.33	-1.14
3. I enjoy research	4.08	1.64	-0.30	-0.99
8. Research is useful for my career	5.59	1.27	-1.21	1.69
9. I find it difficult to understand the concepts of research	3.85	1.59	0.06	-0.99
10. I make many mistakes in research	4.00	1.36	0.04	-0.30
13. I am interested in research	4.59	1.47	-0.57	-0.44
14. Research is connected to my field of study	5.53	1.30	-1.14	1.10
15. Most students benefit from research	5.29	1.13	-0.63	0.47
17. Research is very valuable	5.85	0.98	-1.00	1.24
20. The skills I have acquired in research will be helpful to me in the future	5.46	1.26	-1.16	1.48
21. Research is useful to every professional	5.38	1.26	-0.87	0.26
22. Knowledge from research is as useful as writing	4.78	1.41	-0.49	-0.18
24. Research should be indispensable in my professional training	4.88	1.28	-0.32	-0.43
27. I will employ research approaches in my profession	5.24	1.25	-1.07	1.27
28. Research is difficult	4.90	1.44	-0.77	0.01
29. I am inclined to study the details of research	4.03	1.62	-0.18	-1.03
30. Research acquired knowledge is as useful as arithmetic	3.51	1.41	0.09	-0.74
31. Research-oriented thinking plays an important role in everyday life	4.21	1.56	-0.22	-1.03

Results and Discussion

Model Fit Indices

Using MLE, a CFA was employed to specify the three models tested. Various indicators of fit were utilized to examine the multiple aspects that may encompass a model, to determine how closely a model represented the data, and also how multiple models' fit indices compared to one another (cf. Hu & Bentler, 1999). As with guidelines provided for effect size measures (cf. Cohen, 1988), cut-off points used with indicators of the goodness- or the badness-of-fit for a model serve as suggestions and should be understood and evaluated as evidence of model fit, or lack thereof, within the context of other data presented.

In terms of sample data fit of a model's covariance structure to the observed covariance structure, chi square (χ^2) values that are < .05 signify that a model may be a bad fit for the data, whereas χ^2 values > .05 may render the model a good fit due to the two covariance structures not being statistically significantly different from each other. However, the χ^2 statistic as a measure of fit is known to be sensitive to multivariate non-normality (e.g., skewness and kurtosis) and sample size (e.g., small samples < 200 may have χ^2 values that are not statistically significant, which can lead to type II errors and large samples $\geq$ 200 may produce statistically significant χ^2 that can yield type I errors); all of which may cause uncertainty concerning the overall appropriateness of a study's model based on this particular measure (Bentler, 1990; Browne & Mels, 1992; Cheung & Rensvold, 2002; Hu & Bentler, 1995). Because of these caveats, χ^2 values were reported, but not given much weight in terms of model selection analysis. Another χ^2–based test noted in the literature was more relied upon. The ratio of χ^2 to degrees of freedom (χ^2/df

ratio) was used to compare the relative fit of the three models, where as the χ^2/df ratio decreased, the fit of a model was improved (Hoelter, 1983). It has been suggested that a χ^2/df ratio of 2:1 signifies a good fit (Hair et al., 1995; Kline, 1998).

As incremental fit measures, the comparative fit index (CFI) and the Tucker-Lewis index (TLI) were employed. For both indices, the proposed model needed to compare very well to, or exceed, a null model per the cut-off point $\geq$ .90, which indicate reasonable fit of the model (Kline, 1998; Schumacker & Lomax, 1996). Finally, the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) were used as absolute fit measures to assess the models under study. RMSEA scores of .05, .08, and, .10 have been suggested to represent the magnitude of population misfit, where values < .05 indicate close model fit and values between .05 and .08 indicate reasonable error (Browne & Cudeck, 1993; Hu & Bentler, 1998). For the SRMR, desired cut-off values have been noted at the levels of .05, .08, and, .10 as well (Garson, 2008). Typically, SRMR values $\leq$ .08 are preferred for model fit (Hu & Bentler, 1998). Finally, the Expected Cross Validation Index (ECVI) was used as a single sample estimate of cross-validation to assess how well the study's three models "would generalize to other samples …. The model with the smallest ECVI indicates the model with the best fit." (Hoekstra, Bartels, Cath, & Boomsma, 2008, p. 1158).

One-Factor Model

Looking at Table 3, the one-factor model did not meet the suggested cut-off points (i.e., $\geq$ .90) for the CFI (.30) and the TLI (.25). For the rest of the fit indices, this model showed that it was a weak fit. The one-factor model had χ^2 = 2699.70(464, p < .001) and a 5.82 χ^2/df ratio. The χ^2/df ratio was above the suggested 2:1 ratio. The SRMR (.58) and the RMSEA (.15) values were above the suggested cut-off points of $\leq$ .08. The ECVI was 13.930 and significantly higher than the other two models. Given these immoderate results, it was concluded that the one-factor model had invasive problems with model misspecification and did not fit the data.

Five-Factor Model

The five-factor model had χ^2 = 984.59(454, p < .01) and a 2.17 χ^2/df ratio. The χ^2/df ratio was above the suggested 2:1 ratio. Both the CFI (.83) and the TLI (.82) values were below the cut-off of $\geq$ .90. The RMSEA value of .08 was equal to the suggested value of $\leq$ .08. The SRMR value (.16) was above the suggested cut-off point of $\leq$ .08. The ECVI was 5.579 or over 200% higher than the three-factor model's ECVI value. As can be seen in Table 3, these fit indices results suggested pervasive problems with model misspecification for the five-factor solution and indicated that this model did not fit the data adequately.

Three-Factor Model

The three-factor model had χ^2 = 247.05(132, p < .01) and a 1.87 χ^2/df ratio. The χ^2/df ratio was below the suggested 2:1 ratio. Both the CFI (.92) and the TLI (.91) values were above the cut-off of $\geq$ .90. The RMSEA value of .07 was below the suggested value of $\leq$ .08. The ECVI was 1.779 and substantially lower in value than the other two models, which indicated that the three-factor model's degree of cross-validation, or ability to replicate the current sample data to another sample, was more stable due to its lesser extent of model misspecification (Bandalos, 1993). Thus, the results from Table 3 suggested that the three-factor model fit the data acceptably and should be pursued further with the current CFA analysis. Additionally, all three of the latent factors were moderately to highly inter-correlated: factors 1 and 2 (r_{xy} = .35), 2 and 3 (r_{yz} = .57), and 1 and 3 (r_{xz} = .74). The fact that each factor was not markedly correlated with one another, or r_{xx} > .85, suggested discriminant validity evidence where the model's subscales measured three independent elements of attitudes toward research (Kline, 1998). However, the SRMR value of .13 was above the suggested cut-off point of $\leq$ .08. This last result may be an indicator of potential model misspecification, which will be analyzed further.

Standardized Residuals and Modification Indices

Although the three-factor model was a reasonably good estimation of misfit to the population correlation matrix, it did have some error noted previously. Thus, because the model's SRMR of .13 was

Table 3. Fit indices

Model	$\chi^2(df)$	χ^2/df	CFI	TLI	SRMR	RMSEA	(90% CI)	ECVI
One-Factor	2669.70(464)	5.82	0.30	0.25	0.58	0.15	(0.15,0 .16)	13.930
Three-Factor	247.05(132)	1.87	0.92	0.91	0.13	0.07	(0.05, 0.08)	1.779
Five-Factor	984.59(454)	2.17	0.83	0.82	0.16	0.08	(0.07, 0.08)	5.579

Note: χ^2 = chi-square, df = degrees of freedom, χ^2/df = ratio of χ^2 to df, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = standardized root mean square residual, RMSEA = root mean square error of approximation with 90% confidence intervals, ECVI = Expected Cross Validation Index

Table 4. Score reliability for the three-factor model

Factor	Research Use	Negative Attributes	Positive Attributes
# Items	10	4	4
Cronbach's α	0.88	0.81	0.82
90% CI α	(0.85, 0.90)	(0.74, 0.87)	(0.78, 0.86)
α^2	0.77	0.66	0.67

Note: α^2 = variance accounted for by a factor

above the recommended value of $\leq .08$, standardized residuals were reviewed to determine if there was model misspecification (cf. Fairchild & Finney, 2006). The literature has recommendations for cut-off values concerning standardized residual ranging from $> |3|$ to as large as $> |10|$ (Byrne, 1998; Garson, 2008), which may indicate that a model's covariance structure is different from the observed covariance structure and, hence, model misfit. For the three-factor model, there were 0 standardized residuals $> |10|$ and $0 > |3|$. In fact, the largest standardized residual was 2.67. Overall, the data derived from this analysis suggested that out of the 153 possible standardized residuals, 0% were identified as possible explanations for potential model misspecification and a plausible reason why the SRMR value did not meet the cut-off of $\leq .08$.

Another measure for examining model misspecification is MI (modification indices) (cf. Fairchild & Finney, 2006). Parameters with high MIs > 100 have been noted as potential areas for structure misfit leading to inflated model χ^2 values (Garson, 2008). The three-factor model did not have MIs > 100. The largest decrease in model χ^2 = 14.77 was between items 21 (Research is useful to every professional) and 22 (Knowledge from research is as useful as writing). To ameliorate model misfit problems by correlating error terms between items ex post facto that may be contributing to model misspecification, however, has been a cautioned practice in the literature (Jöreskog & Sörbom, 1989; MacCallum, Roznowski, & Necowitz, 1992) and was not performed in this study due to the lack of a theoretical foundation for said application.

Reliability

Score reliability was conducted for the three-factor model with 90% confidence intervals (CI) (cf. Fan & Thompson, 2001). A recommended cut-off value for score reliability for survey research is $\alpha \geq .80$ (Henson, 2001; Nunnally, 1978). Table 4 indicated that all of the ATR's subscales were > .80; research use (α = .88), negative attributes of research (α = .81), and positive attributes of research (α = .82). Further, the score reliability for the subscales signified that there was high internal consistency and the items that comprised each latent factor shared a large percentage of the variance ranging from .66 to .77.

Further, Table 5 shows communalities (h^2), which are the proportion of each item explained by a latent factor (i.e., akin to R^2). The model's communalities ranged from .21 to .73 with a median of .43. The research use latent factor had h^2 values ranging from .32 to .58 with a median of .43. The negative factor had h^2 values ranging from .21 to .60 with a median of .37. The positive factor had h^2 values ranging from .30 to .73 with a median of .61. Hair et al. (1995) recommended .50 as the cut-off point for this variance accounted for measure, which in terms of average was met by one of the model's latent factors (i.e., positive) and closely by a second (i.e., research use). Individually, item 13 (i.e., I am interested in research) had the highest h^2 value = .73 and item 10 had the lowest h^2 = .21; meaning this variable had a substantial amount of standardized error variance affiliated with it that was unexplained by the negative attributes latent factor (i.e., 79% or 1.00 - .21).

Table 5. Three-factor model pattern and structure coefficients and variance accounted for measure

Item	Factor	P	S	h²
	Factor: Research Use			
8. Research is useful for my career		0.62	0.62	0.40
14. Research is connected to my field of study		0.66	0.66	0.43
15. Most students benefit from research		0.69	0.69	0.48
17. Research is very valuable		0.57	0.57	0.32
20. The skills I have acquired in research will be helpful to me in the future		0.76	0.76	0.58
21. Research is useful to every professional		0.66	0.66	0.44
22. Knowledge from research is as useful as writing		0.58	0.58	0.34
24. Research should be indispensable in my professional training		0.65	0.65	0.43
27. I will employ research approaches in my profession		0.64	0.64	0.41
31. Research-oriented thinking plays an important role in everyday life		0.65	0.65	0.43
	Factor: Negative Attributes of Research			
1. Research makes me anxious		0.78	0.78	0.60
9. I find it difficult to understand the concepts of research		0.58	0.58	0.34
10. I make many mistakes in research		0.46*	0.46	0.21
28. Research is difficult		0.63	0.63	0.40
	Factor: Positive Attributes of Research			
3. I enjoy research		0.77	0.77	0.59
13. I am interested in research		0.85	0.85	0.73
29. I am inclined to study the details of research		0.55	0.55	0.30
30. Research acquired knowledge is as useful as arithmetic		0.79	0.79	0.62

Note: P = pattern coefficient, S = structure coefficient, h² = variance accounted for, * = did not reach a coefficient level ≥ .50 for item salience

Conclusions

The CFA conducted suggested that the three-factor model supported the sample data better than the one- or the five-factor structures, and numerous model fit indices comparison results indicated that the three-factor model was the more preferred model. The three-factor model displayed good model fit in all indices examined, except for the SRMR value of .13, which was above the suggested value of ≤ .08. Because the SRMR was above the desired cut-off value, other analyses were conducted to determine if this result was an indication of potential model misspecification. Analyses via standardized residuals, MI, pattern/structure coefficients, and standardized error variance indicated that item 10, related to the negative attributes latent factor, showed some indicators, but not all conducted in the aforementioned analyses, of possible rewriting to align closer with its current factor or represented a fourth, external domain, where more items of its ilk should be added to have factor manifestation (cf. Fairchild & Finney, 2006; Hertel, 2002).

There were limitations affiliated with the current study. The original ATR scale was investigated via a sample of undergraduate students in education from Cyprus. The current model used participants from the U.S. who were graduate students in education. Also, the U.S. sample spanned 17 majors within education, where the Cyprian sample contained two majors. On two levels, there may be some measurement variance between the participants from the two different countries who may have interpreted the scale's items and latent constructs differently, as well as educational level measurement variance between undergraduate and graduate students and possibly by major.

In terms of future research, the use of the ATR scale's potential predictive ability to forecast earlier in the graduate research process students' attitudes toward research may be of assistance to divert possible problems later on during thesis or dissertation work. Further, the ATR scale should be employed to samples from other countries to test for its equivalence in different contexts and student groupings (e.g., undergraduate/graduate, education/humanities) and also to examine its properties of measurement invariance.

References

Arbuckle, J. L. (2006). AMOS (Version 7.0). [Computer software]. Chicago: SPSS.

Baker, R., & Siryk, B. (1999). *SACQ: Student Adaptation to College Questionnaire manual*. Los Angeles: Western Psychological Services.

Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research, 28*, 351-374.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P. M., & Wu, E. J. C. (1993). *EQS/Windows user's guide: Version 4*. Los Angeles: BMDP Statistical Software.

Björkström, M. E., & Hamrin, E. K. F. (2001). Swedish nurses' attitudes towards research and development within nursing. *Journal of Advanced Nursing, 34*, 706–714.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Research, 21*, 205-229.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Thousand Oaks, CA: Sage.

Browne, M. W., & Mels, G. (1992). *RAMONA users guide*. Columbus, OH: The Ohio State University.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

Cashin, S. E., & Elmore, P. B. (1997, March). *Instruments used to assess attitudes toward statistics: A psychometric evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Combs, J. P., Bustamante, R. M., & Wilson, E. D. (2007). Research skills for effective school leadership: School administrator perceptions. In L. K. Lemasters & R. Papa (Eds.), *At the tipping point: Navigating the course of the preparation of educational administrators* (pp. 310-318). Lancaster, PA: Proactive.

Fairchild, A. J., & Finney, S. J. (2006). Investigating validity evidence for the Experiences in Close Relationships-Revised Questionnaire. *Educational and Psychological Measurement, 66*, 116-135.

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An *EPM* guidelines editorial. *Educational and Psychological Measurement, 56*, 1026-1036.

Garson, G. D. (2008). *Multiple regression*. Retrieved January 28, 2008, from http://www2.chass.ncsu.edu/garson/pa765/regress.htm

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.

Henson, R. K., Hull, D. M., & Williams, C. S. (2010). Methodology in our education research culture: Toward a stronger collective quantitative proficiency. *Educational Researcher, 39*, 229-240.

Hertel, J. (2002). College student generational status: Similarities, differences, and factors in college adjustment. *Psychological Records, 52*, 3-18.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*, 1555-1566.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness of fit indices. *Sociological Methods and Research, 11*, 325-344.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55.

Jöreskog, K. G., & Sörbom, D. G. (1989). *LISREL 7 user's reference guide*. Chicago: Scientific Software.

Kennedy, R. L., & McCallister, C. J. (2001). *Attitudes toward advanced and multivariate statistics when using computers*. (ERIC Document Reproduction Service No. ED464097)

Kline, R. B. (1998). *Principals and practice of structural equation modeling*. New York: Guilford.

MacCallum, R., Roznowski, M., & Necowitz, L. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490-504.

Mills, J. D. (2004). Students' attitudes toward statistics: Implications for the future. *College Student Journal, 38*, 349-361.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S, & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*, 430-445.

Murtonen, M. (2005). University students' research orientations: Do negative attitudes exist toward quantitative methods? *Scandinavian Journal of Educational Research, 49*, 263-280.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*, 353-377.

Newsom, J. (2005). *Practical approaches to dealing with nonnormal and categorical variables*. Retrieved February 22, 2008, from http://www.upa.pdx.edu/IOA/newsom/semclass/ho_estimate2.doc

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Onwuegbuzie, A. J. (1997). Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Library and Information Science Research, 19*, 5-33.

Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological Measurement, 63*, 1020-1038.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment and Evaluation in Higher Education, 29*, 3-19.

Onwuegbuzie, A. J., DaRos, D. A., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics, 19*, 11-35.

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Richard & S. G. Rayner (Eds.), *Self-perception* (pp. 239-265). Westport, CT: Ablex Publishing.

Papanastasiou, E. C. (2005). Factor structure of the 'Attitudes Toward Research' scale. *Statistics Education Research Journal, 4*, 16-26.

Ranis, S. H. (2003, April). *Needing to know: Education leadership preparation and research literacy*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Ravid, R., & Leon, M. R. (1995, April). Students' *perceptions of the research component in Master's level teacher education programs* . Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Richardson, D., & Onwuegbuzie, A. J. (2002, November). *Attitudes toward African-American graduate students as a function of locality*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*, 868-875.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In J. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

Tang, T.L.P., & Chamberlain, M. (1997). Attitudes toward research and teaching: Differences between administrators and faculty members. *Journal of Higher Education*, *68*, 212-227.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57*, 5-19.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45*, 401-405.

Send correspondence to: David A.Walker
                            Northern Illinois University
                            Email: dawalker@niu.edu

Selecting Cluster Results by Classification Accuracy

Mary G. Lieberman   **John D. Morris**

Florida Atlantic University

Cross-validated predictive discriminant analysis was used to illustrate comparison of cluster solutions by classifying subjects into candidate clusters. This technique was illustrated for both selection of the number of clusters within a clustering algorithm and the selection of a clustering algorithm. Other applications to demonstrate and judge the adequacy of potential clustering results, and to assess the accuracy of contrasting clustering algorithms, are suggested.

T he purpose of this study was to investigate and demonstrate a method by which one could aid decisions regarding cluster adequacy by using classification accuracy. Examples include decisions regarding cluster number and algorithm, although other uses are proposed. The thorny problem of how many clusters, how large they should be, and who should belong to them in cluster analysis is parallel to the same concerns in factor analysis. Indeed, the origins of cluster analysis in "inverted" or transformed factor analysis (i.e., grouping subjects rather than variables) points to this theoretical similarity. Loosely speaking, cluster analysis is a class of techniques that are used to determine whether subjects tend to group together in respect to their scores on variables in multivariate space (see Anderberg, 1973 for an introduction, but distinguished from early factor usage by Harman, 1967). A distinction is necessary between the cluster analysis process of creating groups of subjects based on their statistical "likeness," and the process of classifying subjects into groups that are known *a priori* in what is typically called predictive discriminant analysis (viz., Huberty, 1994 for a complete description); or classification analysis; or, switching mathematical models, logistic regression. In the former method, the objective is to decide if there is any defensible "natural" grouping of subjects in respect to the subjects' score vectors; in the latter procedure, interest is in ascertaining the accuracy with which one can classify subjects *a priori* into known groups conditional on their score vectors. If there is such a grouping, or "clustering," the researcher must decide on the number of clusters, the size of the clusters, and who belongs to those clusters. As in factor analysis, a variety of quantitative aids have been developed to help make decisions regarding the clusters resultant from a cluster analysis, but also, as could be argued in factor analysis, the decision is still primarily subjective. That is, the aids merely offer quantitative indices about which subjective decision can be made.

We proffer a more objective, and what we believe to be more meaningful, criterion of the merit of any particular clustering result, and also an inferential way of comparing the merit of candidate clustering results in respect to this criterion. So, the thrust herein is not primarily about candidate cluster analyses, nor a comparison among clustering algorithms, but regards a recommendation for a criterion to be used in judging the of adequacy of cluster solutions, from whatever algorithm and dimensionality decision, and for a method of comparing alternate clustering solutions on this criterion.

By definition, clustering algorithms function based upon some variant of the notion of similarity, or distance, calculated between each case and every other case, and then grouping the cases that have the greatest similarity, or the least distance, into clusters. Differing results obtained based on the type of distance or similarity indices used (i.e., correlation or a Euclidian or non-Euclidean scale preserving metric), the clustering algorithm, and even holding these constant, upon decisions regarding such issues as number of clusters. Thus, classification can aid in this decision (see Huberty, DiStefano, & Kamphaus, 1997 and Lieberman, Morris, & Huberty, 2001 for a description of the steps in a cluster analysis and the use of classification, and other methods, in the validation of clusters).

Method

What we propose is to judge the adequacy of any potential clustering result by the accuracy with which one can classify the subjects into the groups derived by the cluster analysis. Such classification can be accomplished via predictive discriminant analysis, logistic regression, or other methods. Therefore, in this case, one is treating the groups derived from the cluster analysis as if they were "known." This can be done for any clustering solution as a validation step, but, as well, candidate clustering solutions can be compared -- whether it is simply a question of how many clusters one should retain within a clustering algorithm, across different clustering algorithms, both, or other differing clustering method attributes. Candidate cluster results can be compared on the classification accuracy obtained across all groups (or

clusters), or by separate groups (or clusters), though the total group accuracy seems of most relevance. These accuracies can be compared for classification of the calibration sample, or for cross-validated estimates of classification accuracy, as in the "leave-one-out" procedure that is popular in predictive discriminant analysis (Huberty, 1994). Moreover, one could elect to use a linear classification function (assuming equal group – or cluster – covariance matrices), or a quadratic function (where covariance matrices are not assumed equal). Prior probabilities are considered to be of no relevance in this application in which the "groups" are the product of a cluster analysis. Furthermore, the uncritical use of unequal prior probabilities has been questioned (Meshbane & Morris, 1995a). Nonetheless, this stand is potentially debatable and the algorithm easily accommodates use of unequal prior probabilities if the researcher wishes. Moreover, the relative accuracy resulting from the different clustering structures can be contrasted inferentially using McNemar's contrast for correlated proportions in the same way as has been suggested to compare full and reduced classification models in predictive discriminant analysis (Morris & Huberty, 1995a) and linear and quadratic models (Meshbane & Morris, 1995b).

Results and Discussion
Example 1: Comparison of Number of Clusters
An example data set, taken from George and Mallery (1999), was used to illustrate the method herein. The research question addressed in that example involved clustering 21 VCRs (videocassette recorders) based on a variety of characteristics including: price, an index of picture quality, an index of programmability, two indices of reception, three audio reception indices, three indices regarding remote quality, indices regarding events, days of programmability, features, and "extras." In this case, the method selected by the authors was a hierarchical/agglomerative cluster analysis using Euclidean distances. Perusal of the results led the authors to a subjective decision that three clusters most adequately represented the data.

To illustrate the method that we propose, we analyzed the same data with the same procedure using the same software (Statistical Package for the Social Sciences), but selected a two cluster solution. We then performed a leave-one-out type cross-validated discriminant analysis classifying the subjects alternatively into the two- or three- cluster solutions. We then contrasted the total-group hit rate for the two solutions using the McNemar statistic. The McNemar contrast for correlated proportions requires a joint frequency distribution; as applied to this situation, a joint frequency distribution of hits and misses for the two-cluster and three-cluster solutions. Table 1 shows that, out of 21 cases, 6 were correctly classified by the two cluster solution that were incorrectly classified by the original three cluster solution with 15 subjects correctly classified by both solutions. Both the z, or exact binomial test, could be considered statistically significant ($p < .05$) in this case, with the alternate two-cluster arrangement manifesting superior classification results. Increased parsimony and accuracy was achieved in respect to the originally posited three cluster solution.

Table 1. Contrast of two and three cluster classification performance

		Three Cluster		
		Miss	Hit	
Two Cluster	Hit	6	15	21
	Miss	0	0	0
		6	15	

McNemar $z = 2.45$ $p = 0.014$
Exact Binomial $p = 0.031$

Example 2: Comparison of Clustering Algorithms
As was mentioned, an alternative application of the method is to judge the adequacy of the clustering algorithm based on classification accuracy. For this same data set, a Two-Step and k-Means clustering algorithm (both with two groups) were compared (see Table 2). Note, that although two groups were used in both algorithms, different cases belonged to the clusters according to the result of the clustering algorithm. One can see that the k-Means algorithm provided clusters for which classification accuracy might be considered statistically significantly ($p < .05$) more accurate than that from the Two-Step algorithm.

Table 2. Contrast of two-step and k-means classification performance

		Two-Step		
		Miss	Hit	
k-Means	Hit	7	14	21
	Miss	0	0	0
		7	14	

McNemar $z = 2.65$ $p = 0.008$
Exact Binomial $p = 0.016$

Conclusion

We suggest that serious consideration be given to the criterion of classification accuracy afforded by a cluster solution in decisions regarding cluster results. Moreover, the technique is equally applicable in contrasting the cluster results (if different) from other clustering algorithms, the same algorithm with different variables (in this case there are interesting possibilities for addressing full versus restricted model questions), or any of a variety of methodological or subjective alternatives that give different cluster results.

References

Anderberg , M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

George, D., & Mallery, P. (1999). *SPSS for Windows step by step: A simple guide and Reference*. New York: Allyn & Bacon.

Harman, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.

Huberty, C. J. (1994). *Applied discriminant analysis*. New York: Wiley.

Huberty, C. J, DiStefano, C., & Kamphaus, R. W. (1997). Behavioral clustering of school children. *Multivariate Behavioral Research, 32*, 105-134.

Lieberman, M. G., Morris, J. D., & Huberty, C. J (2001, April). *Selecting cluster results by classification accuracy*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Meshbane, A., & Morris, J. D. (1995a, April). *Assuming equal vs. unequal prior probabilities of group membership in discriminant analysis: Effect on predictive accuracy*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Meshbane, A., & Morris, J. D. (1995b). A method for selecting between linear and quadratic classification models in discriminant analysis. *Journal of Experimental Education, 63*, 263-273.

Morris, J. D., & Huberty, C. J. (1995). Full versus restricted model testing in predictive discriminant analysis. *Journal of Experimental Education, 63*, 161-165.

Send correspondence to: Mary G. Lieberman
                           Florida Atlantic University
                           Email: mlieberm@fau.edu

Multiple Linear Regression Viewpoints

Multiple Linear Regression Viewpoints (*MLRV*) is a publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model (SIG/MLR: GLM). It is published twice a year to facilitate communication among professionals who focus their research on the theory, application, or teaching of multiple linear regression models and/or the general linear model. Manuscripts submitted to *MLRV* should conform to the language, style, and format of the *Publication Manual of the American Psychological Association* (6th ed., 2009). Three copies (two blind) of a doubled-spaced manuscript (including equations, footnotes, quotes, and references) of approximately 25 pages in length, a 100 word abstract, and an electronic copy with the manuscript formatted in WordPerfect or Word should be submitted to the editor. Mathematical and Greek symbols should be clear and concise. All figures and diagrams must be photo copy ready for publication.

Manuscripts will be anonymously peer-reviewed by two editorial board members. Author identifying information should appear on the title page of only one submitted manuscript. The review process will take approximately 2 to 3 months. A letter acknowledging receipt of the manuscript will be sent to the first author, and upon review, a letter indicating the peer-review decision will be sent to the first author. Potential authors are encouraged to contact the editors to discuss ideas for contributions or determine if their manuscript is suitable for publication in *MLRV*.