The Use of Multilevel Modeling to Estimate Which Measures are Most Influential in Determining an Institution's Placement in Carnegie's New Doctoral/Research University Classification Schema

By:

Theodore Micceri, Ph.D.
Coordinator of Statistical Research
Office of Planning and Analysis

Office of the Provost
University of South Florida

Tampa, FL 33620

April 2007

Table of Contents

Executive Summary

This research sought to determine whether any measure(s) used in the Carnegie Foundation's classification of Doctoral/Research Universities contribute to a greater degree than other measures to final rank placement.

Methods

Multilevel Modeling (MLM) was applied to all eight of the Carnegie Foundation's predictor measures using final rank (doctoral/research, high research, very high research) as the outcome (dependent) variable. Data came directly from the Carnegie Foundation. One additional variable, private or public control came from the IPEDS Peer Analysis System.

Findings

- All measures used in the Carnegie Foundation's analyses exhibited strong interrelationships (multicollinearity), which reduces the reliability of multivariate analyses (Table 1, Table 3).
- The overall MLM regression model predicted approximately 50% of the variance in rank, with an estimated multiple ***r*** of .72.
- The most powerful predictor of rank was federal science & engineering (S&E) expenditures. Once this variable entered the prediction model, only doctorates granted in the humanities added significantly to prediction (3.5% of variance, Table 1).
- Although both the number of post doctoral appointments (.86) and non-faculty researchers (.67) exhibit strong simple relationship with rank (Table 3) when S&E expenditures and humanities doctorates are entered into the MLM model, post doctoral appointments and non-faculty researchers both contributed non-significantly and negatively to predicting Carnegie rank.
- Table 3 shows that most relationships between predictor variables and the outcome Carnegie rank ranged between .75 (number of faculty) and .89 (S&E expenditures). All of the measures also exhibit strong relationships with other predictors. That the number of faculty has a simple Rranks of .75 indicates that a research institution's size alone relates to their rank.

Conclusion

Using eight predictor measures, all of which interrelate strongly and significantly (Table 3), is effectively like using a single measure to rank institutions. An institution's S&E expenditures may thus be effectively used as that single predictor, although doctorates in the humanities can also influence an institution's rank.

The Use of Multilevel Modeling to Estimate Which Measures are Most Influential in Determining an Institution's Placement in Carnegie's New Doctoral/Research University Classification Schema

Purpose

This research sought to determine whether any specific variable(s) used in the Carnegie Foundation's classification of Doctoral/Research Universities contributes more than others to an institution's final placement. Due to its robustness and ability to estimate the effects of independent single unit increments for predictor variables on outcome variables (see Appendix A), this study used Multilevel Modeling (MLM).

Background – The New Carnegie Doctoral/Research Classification Schema

The new system classifies doctoral/research institutions into one of three ranks based on two indices developed from eight measures.

Research Universities (very high research activity)
Research Universities (high research activity)
Doctoral/Research Universities

In the new classification system, USF and Miami were originally designated as high research rather than very high research activity institutions, while FSU and UF were classified as very high. The eight measures upon which the classification rests are:

1. doctoral conferrals in humanities fields
2. doctoral conferrals in social science fields
3. doctoral conferrals in fields other than science, engineering, technology, and mathematics
4. doctorates granted in STEM disciplines
5. postdoctoral appointees
6. non-faculty research staff
7. research & development (R&D) expenditures in science and engineering (S&E)
8. R&D expenditures in non-S&E fields

In the final revision, STEM doctorates were also included in the modeling process.[1]

The primary difference between the former and new classifications are that the old one grouped institutions using only total research expenditures combined with a threshold for total doctorates granted. This put considerable emphasis on the National Science Foundation (NSF) orientation toward supporting research in what are known as STEM (science, technology, engineering, mathematics) and S&E (science & engineering).

The new method places less emphasis on STEM. In their description of the new method, Carnegie notes: "...the funding and staffing variables are sufficient measures of research activity in STEM fields." Four of the seven measures continue to deal primarily with STEM disciplines: (2) social sciences are defined as STEM by NSF, (5) postdoctoral

[1] A senior Carnegie Research Scholar informed the present author that this decision resulted at least partly from his rather ardent suggestions regarding why STEM doctorates should be included following Carnegie's request for feedback to the initial methodology.

appointees are only counted in Science, Engineering and Health fields, (7) R&D in S&E, and, (6) non-faculty research staff, are also probably are more common in STEM than NonSTEM disciplines.

A second difference is the use of two indices rather than a single one. The two are (1) aggregate totals, and (2) aggregate totals standardized by the number of full-time faculty engaged primarily in instruction or research and, instruction, research and public service. The first method should work to the advantage of large institutions (SUS institutions include the five largest in the southeastern U.S.), while the second should work to the advantage of more efficient institutions (those who produce more with the same number of faculty).

Methods

Data Sources

All data derived either from the Carnegie Foundation sources that were used for their analyses in December 2005, or from the IPEDS Peer Analysis System (IPEDSAS).

Analyses were conducted using PC SAS, 9.1, PC SPSS 15.1 and HLM 6 and were summarized using Microsoft Excel.

Because two different forms of predictor variables were used in Carnegie's analyses, one using aggregate numbers, the other using numbers standardized by the number of faculty at an institution, the initial design of this study was to conduct two MLM analyses, once using per capita standardized predictors and once using aggregate predictors. However, the results of the first analysis indicated that the second was unlikely to add useful information. Therefore, only a single analysis, using aggregate data, was conducted.

Variables

The outcome or dependent variable was the Carnegie rank, coded as 3 for Very High, 2 for High and 1 for Doctoral/Research.

The eight predictor variables used by the Carnegie Foundation were included using the values (whether accurate or not) that were used within the Carnegie Analyses.

1. doctoral conferrals in humanities fields
2. doctoral conferrals in social science fields
3. doctoral conferrals in fields other than science, engineering, technology, and mathematics
4. doctorates granted in STEM disciplines
5. postdoctoral appointees
6. non-faculty research staff
7. research & development (R&D) expenditures in science and engineering (S&E)
8. R&D expenditures in non-S&E fields

In addition to these, an additional Level II predictor was added in these analyses, institutional control, where private institutions received a 1 and public, a 0 (zero).

MLM Nesting

Institutions were nested within states, which assured that adequate k-level variables were available for the MLM analyses to have power to detect differences where they occurred.

Limitations

Because only 256 doctoral/research institutions were used by the Carnegie Foundation, and because nine predictor variables were included in the analyses, it does not appear legitimate to conduct an internal cross validation to verify the outcomes due to sample size limitations. Due to the numerous tests that are used in MLM models, only effects significant at the $\boldsymbol{p} < .001$ level are treated as significant.

Results

Sample

The sample included 256 Doctoral/Research universities in 52 states (Puerto Rico and the District of Columbia included). The number of institutions per state ranges from 22 in California, to one in eight states. The average number of institutions per state is 4.9.

Multilevel Modeling Analysis

MLM analysis, using HLM 6 was applied to all predictive variables used by the Carnegie Foundation in their ranking process (Table 1). Although the value distributions for almost all of these variables included both extreme asymmetry and outliers, as is explained in Appendix A, MLM is not as influenced by these distributional contaminations as are traditional Ordinary Least Squares (OLS) forms of analysis. This effect is explained for Table 3, which also demonstrates the high degree of multicollinearity among these predictors. This effect also explains the extremely low reliability obtained in the MLM predictive process (bottom row of Table 1).

Overall, the MLM models developed using these several predictors account for roughly 50% of the variance in the outcome variable Carnegie Rank ($\boldsymbol{r} = .70$). Due to the multicollinearity, the model including only one predictor (in this case, S&E Expenditures – Model I) produced approximately the same level of unexplained variance in the model as Model V, which included all predictors. The Level II variable, whether an institution is public or private, added nothing to the explanatory power of the model.

Once S&E Expenditures was in the model, only three additional variables added any predictive capacity to the overall model, and in each case these were small additions. The largest was doctorates in the humanities, which added 3.2% to the variance predicted. Additionally, social sciences doctorates and non-faculty researchers contributed small amounts. Although their simply correlations with outcomes was positive (Table 2), in the MLM model, with other predictors entered, both non-faculty researchers and post doctorates added negative contributions to the prediction (more of either means lower ranks).

Finally, to summarize the effects of different measure values on outcomes, the lowest values produced an average rank between Doctoral/Research and High Research, average values produced a rank slightly above High Research, and maximum values produced a rank well above Very High Research.

Table 1
Multilevel Modeling Results with Carnegie Rank as Outcome (N=256)

	Model 0	Model I	Model II	Model III	Model IV	Model V
Level 1						
Intercept	2.1	1.7	1.7	1.7	1.7	1.7
t	43.2***	42.2***	41.4***	40.3***	34.3***	30.7***
S&E_Expends		0.000003	0.000003	0.000003	0.000003	0.000002
t		14.2***	10.5***	10.1***	8.2***	4.2***
Docs Humanties			0.005003	0.005180	0.005158	0.001064
t			3.9***	3.9***	3.1**	0.4
Non-Significant Effects						
PostDocs				-0.000130	-0.000130	-0.000074
t				1.3	-0.9	-0.5
Non S&E Expn						0.000001
t						0.2
NonFac Rschrs						-0.001985
t						-2.1*
Docs Sociology						0.007605
t						2.4*
Docs Professnal						-0.000987
t						0.0
Docs STEM						0.001939
t						1.9
Level II						
Public/Private					0.0856	0.032633
t					0.32	0.12
Variance Explained by a Given Model						
Model to Model Change		48.8%	3.4%	0.0%	-0.2%	5.2%
***r* and *R*^{2} Estimates**						
r	0.00	0.69	0.70	0.59	0.70	0.72
***R*^{2}**		0.47	0.49	0.34	0.49	0.51
	Model Reliability					
Intercept	0.00	0.12	0.12	0.13	0.14	0.15

* ***p*** < .05
** ***p*** < .01
*** ***p*** < .001

Within State Results

Despite its lack of value from a predictive perspective, providing no gain over a fixed effects model, a random effects model, with variable slopes and intercepts was used to provide separate regression models for each of the 21 states having at least four doctoral/research institutions (Table 2). This was done to see what differences occur within states in the relationship between S&E Expenditures and Carnegie rank. Note that the mathematical form of this predictive model is:

$$Y = a + b_1 x_1$$

Where:

Y	= predicted Carnegie rank
a	= Y-axis intercept
b_1	= slope coefficient
x_1	= research dollars

In Table 2, $\mathbf{R^2}$ shows the amount of variance in the outcome variable (Carnegie Rank) that is accounted for by the prediction equation. The square root of this value shows the simple correlation. Thus, even what appears to be a low $\mathbf{R^2}$ of .36 (Massachusetts) represents a fairly strong relationship (***r*** = .60) between the predictor (Federal Research) and the outcome variable (Carnegie Rank).

Table 2
Within State Regression Model of Federal S&E Research with Carnegie Rank

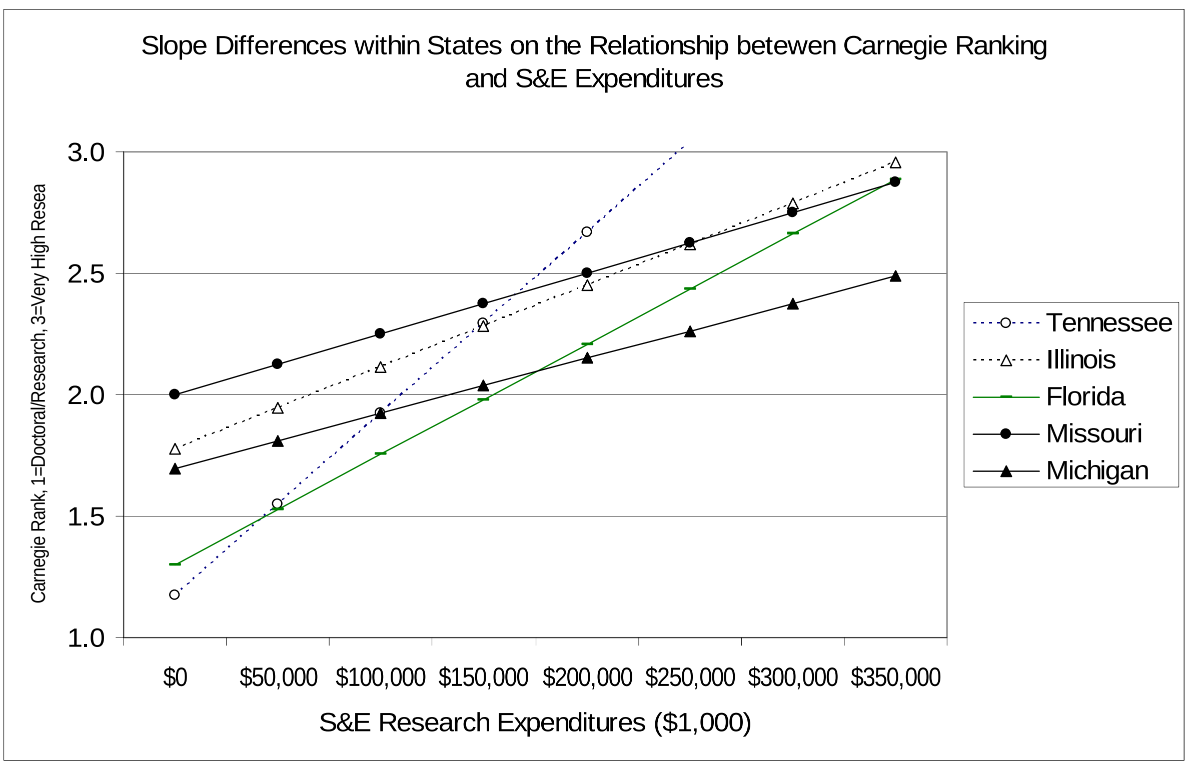
State	N	$\mathbf{R^2}$	Intercept	Slope[2]
California	22	0.53	1.55	0.00000270
Colorado	5	0.85	1.46	0.00000743
District of Columbia	5	0.65	1.32	0.00001325
Florida	12	0.59	1.30	0.00000454
Georgia	5	0.89	1.37	0.00000498
Illinois	9	0.72	1.77	0.00000338
Indiana	6	0.40	1.56	0.00000513
Louisiana	5	0.71	1.48	0.00000878
Massachusetts	11	0.36	1.96	0.00000292
Michigan	7	0.52	1.70	0.00000227
Minnesota	5	0.99	1.00	0.00000393
Missouri	6	0.78	2.00	0.00000251
North Carolina	7	0.72	1.69	0.00000308
New Jersey	6	0.78	1.57	0.00000659
New York	21	0.51	1.77	0.00000363
Ohio	14	0.47	1.58	0.00000393
Oregon	4	0.90	1.04	0.00001214
Pennsylvania	11	0.72	1.42	0.00000353
Tennessee	5	0.86	1.17	0.00000749
Texas	15	0.47	1.59	0.00000376
Virginia	6	0.85	1.74	0.00000511

Figure 1 depicts the simple within-state regression between Carnegie ranks (where 3=Very High Research and 1=Doctoral/Research) and federal S&E research expenditures. As Table 1 and Table 3 show, this is the "best" single predictor of rank,

[2] These coefficients are small because Federal Research Expenditures are $ hundreds of thousands.

although other factors such as doctorates in the humanities, STEM doctorates and the number of faculty also have some influence. The figure shows a few examples of the differing slopes that associate with institutions in specific states having at least five doctoral/research universities. Part of the slope and intercept effect (where $0 intercepts the Y-Axis or Carnegie Rank) relates to how many institutions in a state are at a given rank. For example, Tennessee, with five institutions, shows the steepest slope,[3] but this occurs because Tennessee has two institutions at the lowest rank, Doctoral Research (East Tennessee State University and Tennessee State University), one at High Research (University of Memphis) and two at Very High Research (Vanderbilt University and the University of Tennessee). Missouri, with six institutions, shows the least steep slope, because their lowest ranked institutions are all High Research (University of Missouri Kansas City, Rolla, Saint Louis, and Saint Louis University), and they have two at Very High Research (University of Missouri-Columbia and Washington University-St. Louis). Florida, with 12 institutions, exhibits a comparatively steep slope because there are five at Doctoral/Research (Barry University, FAMU, FAU, Nova Southeastern, UWF), three at High Research (UCF, FIU, FIT-Melbourne), and four at Very High Research (UF, FSU, USF, University of Miami). Thus, states having the entire range on Carnegie rank tend to exhibit steeper slopes than those having only part of the possible range on the outcome variable Carnegie rank.

Figure 1



[3] The District of Columbia (DC), also with five institutions has a steeper slope, but was too extreme to well depict the between state variability for this chart.

Table 3 shows simple Spearman Rranks correlations between and among the variables used by the Carnegie Foundation to establish Doctoral/Research ranks. Rranks was used rather than Pearson's r due to the extremely great asymmetry present in many of these variables as well as the several extreme outliers and leverage points. As an example, for post doctorates, Harvard had 3,698, while the second greatest was 1,351 at UCLA. Such extremes frequently create rather severe under or over-estimates for Ordinary Least Square (OLS) statistics like $\boldsymbol{r}$. Therefore, robust rank statistics (Spearman) were used to produce reasonably reliable estimates. This doesn't cause much of a problem for MLM estimates, as is explained in Appendix A.

All of the relationships shown in Table 3 are significant at the $\boldsymbol{p} < .0001$ level. The fact that every variable has a relatively strong relationship with every other variable in the list shows the multicollinearity present among these variables. This is also evidenced by the very low reliability of the MLM predictive model in Table 1. Such strong relationships among all variables demonstrate that there is basically one factor/predictor at work in these rankings. The first column, relationships with rank show that S&E expenditures, post doctorates, STEM doctorates, total doctorates and the number of faculty all predict roughly half or more of the variance in the outcome variable rank (High Research, Very High, etc.).

Table 3

Spearman Rank Correlations Between and Among Variables Used by Carnegie to Create Ranks (N=256)

						Doctorates Granted				
	Carnegie Level	S&E Expenditures	Non-S&E Expenditures	Post Doctorates	Non-Faculty Researchers	All Fields	Humanities	Social Sciences	Professional Fields	STEM
S&E Expenditures	0.89									
Non-S&E Expenditures	0.54	0.59								
Post Doctorates	0.86	0.87	0.49							
Non-Faculty Researchers	0.67	0.68	0.37	0.77						
Total Doctorates	0.77	0.77	0.53	0.76	0.58					
Humanities Doctorates	0.66	0.62	0.54	0.65	0.51	0.78				
Social Sciences Doctorates	0.70	0.65	0.52	0.66	0.53	0.81	0.78			
Professional Doctorates	0.37	0.43	0.37	0.41	0.33	0.75	0.50	0.49		
STEM Doctorates	0.88	0.91	0.54	0.87	0.68	0.85	0.66	0.70	0.43	
Number of Faculty	0.75	0.86	0.62	0.80	0.63	0.78	0.70	0.70	0.56	0.80

Analysis of Residuals

Although it is usually wise in MLM analyses to evaluate how badly assumptions are violated, due partly to the multicollinearity and consistent nature of findings in this study (e.g. Table 3), and partly to the extreme distributional contaminations present in most of the measures, no analysis of residuals was conducted because assumptions are violated so severely that such analyses are probably not very meaningful.

Summary and Discussion

The findings of this study indicate:

- All measures used in the Carnegie Foundation's analyses exhibited strong interrelationships (multicollinearity), which reduces the reliability of multivariate analyses (Table 1, Table 3).
- The overall MLM regression model predicted approximately 50% of the variance in rank, with an estimated multiple ***r*** of .72.
- The most powerful predictor of rank was federal science & engineering (S&E) expenditures. Once this variable entered the prediction model, only doctorates granted in the humanities added significantly to prediction (3.5% of variance, Table 1).
- Although both the number of post doctoral appointments (.86) and non-faculty researchers (.67) exhibit strong simple relationship with rank (Table 3) when S&E expenditures and humanities doctorates are entered into the MLM model, post doctoral appointments and non-faculty researchers both contributed non-significantly and negatively to predicting Carnegie rank.
- Table 3 shows that most relationships between predictor variables and the outcome Carnegie rank ranged between .75 (number of faculty) and .89 (S&E expenditures). All of the measures also exhibit strong relationships with other predictors. That the number of faculty has a simple Rranks of .75 indicates that a research institution's size alone relates to their rank.

Using seven predictor measures, all of which interrelate strongly and significantly (Table 3), is effectively like using a single measure to rank institutions. An institution's S&E expenditures may be used as that single predictor, although doctorates in the humanities can also influence an institution's rank.

References

Raudenbush & Byrk (2004). *Hierarchical Linear Modeling: Applications and Data Analysis Methods*. Sage Publications, Thousand Oaks, CA.

Bhaskar, R. (1989). *The Possibilities of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. Atlantic Highlands, New Jersey: Humanities Press.

Hox, J. (2002). *Multilevel Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kreft, de Leeuw & Aiken (1995). The Effect of Different Forms of Centering in Hierarchical Linear Models, *Multivariate Behavioral Research, 30* (1), 1-2 1).

Luke, D. A. (2004). *Multilevel Modeling*. Newbury Park: Sage Publications.

Robinson, W.S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15: 350-357.

Appendix A – Reasons for Using Multilevel Modeling Rather than OLS Statistics

Multilevel Modeling (called by many different names), has a lot to recommend it, including, but not limited to:

- First, it can be used for a very wide variety of different purposes, and all with far fewer and less restrictive assumptions than OLS. It can be used in place of the following, just to name a few: meta analysis, repeated measures analysis, multiple regression, and logistic regression.
- Predictive statistical methods (Regression, Hotelling's T, etc.) frequently provide useful information for IR purposes. Multilevel Modeling is rooted in Multiple Regression, but rests upon far less restrictive assumptions than the OLS method, and can produce reasonably accurate predictive models. Also, because Maximum Likelihood (ML) estimates are used to estimate both fixed and random effects, and because test statistics based on robust variance estimates are included in HLM 6 outputs, and because covariance estimates derive from Empirical Bayes (EB) residuals, the validity of output is surely more robust to violations of the underlying normality assumptions than are OLS estimates.
- The predictive models created within the constraints implemented when using this method allow one to estimate specific influences on outcomes from incremental increases or decreases in predictor variables in ways that are more meaningful and useful than those typically developed by using traditional regression models.
- Almost all who teach or write about Multilevel Modeling and any skilled individual who is teaching about Regression, will make it a point to emphasize how important it is to check the underlying distributional assumptions, and particularly, to conduct analysis of Residuals' distributional characteristics.
- Most who use Multilevel Modeling use graphs/charts to better see what's occurring. A picture is worth 1,000 words.
- Multilevel modeling allows one to readily estimate Intraclass Correlation Coefficients (ICC), which can substantially alter error terms in analyses and thereby create false positives. A common rule of thumb is to use multilevel modeling when ICC is greater than 0.05.
- MLM treats the open systems in which research is almost always conducted as open, rather than closed. The traditional Experimental Paradigm is based upon 19th Century, reductionist, closed system thinking, which really only applies to a few things like the movement of planetary bodies, and to absolutely nothing in the Social and Behavioral Sciences. Because everything nests within a larger context, it makes sense to use a technique for analysis that can take this nesting into account.
- Following on the preceding, Multilevel Modeling nests dependent variables within contexts which exert influence on those dependent variables in the real world. A summary of Kreft, de Leeuw & Aiken (1995) says it well: "In multilevel models, micro-level units, such as workers or students, are nested within macro-

level units, such as industries or schools. In multilevel models, separate predictors characterize the micro-level units, the *individuals,* and the macro-level units, the *groups* or *contexts*. The assumptions regarding the coefficients of the model depend upon the level of the predictors. The coefficients of all but the highest level predictors may be treated as random; hence the name *random coefficient models,* while those of the highest level are always treated as fixed."

- The process of building Multilevel Models makes sense from a research perspective because you first construct an unconditional model (like a ONEWAY ANOVA), determine whether it makes sense to use Multilevel Modeling given the nature of the data relative to the question, and advance step-by-step through the process of model development to a final point, checking in multiple ways at each step to see whether what you are doing is a good idea or not, until you finally have a full model, which, hopefully, provides a reasonably accurate prediction of your dependent variable.

- Because all social science contexts are complex, only analyses that can isolate the unique impact (unique variation) of specific factors at their various levels, such as multilevel modeling, are appropriate. Effectively, Multilevel Modeling uses Backward Elimination Regression rather than Stepwise to model equations thereby identifying only the unique contribution of each variable to a model.

- Multilevel modeling can allow one to develop a regression (prediction) model for each context separately rather than assuming that a single average prediction model applies in all groups.

The Simplest Arguments for Using Multilevel Modeling

MLM is widely applicable in situations that many wish to analyze, as Raudenbush & Byrk (2002, p. 142) note: "One of the most common applications of HLM in organizational research is simply to estimate the association between a level-II predictor (*in this case, sex*) and the mean of Y, adjusting for one or more level-I covariates."

Luke (2004, p. 7) makes the simplest argument:

> The simplest argument, then, for multilevel modeling techniques is this: Because so much of what we study is multilevel in nature, we should use theories and analytic techniques that are also multilevel. If we do not do this, we can run into serious problems like the Ecological Fallacy. Where relationships observed in groups are assumed to hold for individuals (Robinson, 1950). The Atomistic Fallacy, in which inferences about groups are incorrectly drawn from individual-level information (Hox, 2002)...these fallacies are a problem of inference, not of measurement. That is, it is perfectly admissible to characterize higher-level collective using information obtained from lower-level members. The types of fallacies described above come about when relationships discovered at one particular level are inappropriately assumed to occur in the same fashion at some other (higher or lower) level.