Zen and the Art of Policy Analysis:

A Response to Nielsen and Wolf

Kenneth J. Meier, Texas A&M University

Warren S. Eller, Texas A&M University

Robert D. Wrinkle, University of Texas Pan American

J.L. Polinard, University of Texas Pan American

September 7, 2000

Accepted for Publication, Journal of Politics, May 2001

We thank George A. Krause and Jeff Gill for helpful comments. All statistical analyses were conducted in STATA and replicated in limdep. All data and documentation to replicate this analysis including the complete STATA output can be found at http://web.polmeth.ufl.edu/ or www-bushschool.tamu.edu/pubman/

Zen and the Art of Policy Analysis:

A Response to Nielsen and Wolf

Abstract

Neilsen and Wolf (N.d.) lodge several criticism of Meier, Wrinkle and Polinard (1999). Although most of the criticisms deal with tangential issues rather than our core argument, their criticisms are flawed by misguided estimation strategies, erroneous results, and an inattention to existing theory and scholarship. Our re-analysis of their work demonstrates these problems and presents even stronger evidence for our initial conclusion–both minority and Anglo students perform better in schools with more minority teachers.

Zen and the Art of Policy Analysis:

A Response to Nielsen and Wolf

Policy analysis is an art not a rote exercise. It requires informed judgement at every step of the process and an understanding of the implications of methodological choices. This exchange illustrates the importance of that art. In "Representative Bureaucracy and Distributional Equity: Addressing the Hard Question" we presented two key findings–in school districts with more minority teachers both minority and Anglo students perform better. We also addressed several interesting but tangential issues including the anomaly of overall test scores, the linearity of the relationships, and how equity and performance were linked. Nielsen and Wolf (N.d.) lodge six specific criticisms of our work:

1. The appropriate estimation technique is a fixed effects pooled regression.
2. We omitted a key variable, percent Anglo students, in our model.
3. The relationship between minority teachers and student performance is not nonlinear.
4. Minorities should be disaggregated into blacks and Latinos; when this is done, there is no relationship between minority teachers and minority student performance.
5. The appropriate unit of analysis is the student not the school district.
6. Substantively weighted least squares (SWLS) is used incorrectly.

Even if Nielsen and Wolf were correct (and they are not), only one of these criticisms bear on our key findings; the others focus on peripheral issues. Our response demonstrates that Nielsen and Wolf are incorrect in assertions 1 through 5 and misinterpret assertion 6 because they failed to read the relevant work in this area. We deal with each criticism in order.

1. The Use of Fixed Effects Models

Nielsen and Wolf contend that we have underspecified models, specifically that based on a Hausman specification test we should have estimated our equations as fixed effects models by including a series of dummy variables to represent the unique characteristics of each individual school district. With a panel data set of dimensions N x T (where N is the number of unique cases and T is the number of time periods), there are a myriad of possible models (Baltagi 1995; Hsiao 1986; Greene 1997). The appropriate model depends on the characteristics of the data and any special problems therein (serial correlation, heteroscedasticity, cross-correlations, etc.). Meier et al. (1999) include time dummy variables in all models; the point of contention is whether or not they should have included N-1 cross-sectional dummy variables, a fixed effects or least squares dummy variables (LSDV) model (Greene 1997, 616). Fixed effects models are a possible choice in any N x T panel; but as T approaches 1, the costs of the model often exceed its benefits. At T=1, a fixed effects model is impossible, at T=2 it boggles the mind. At T=6, the current case, the practice is questionable in general because the number of time periods is too few to provide efficient estimation of the intercepts as unique contributions. With the current data set, moreover, fixed effects is a exceptionally poor choice for two reasons–the lack of remaining variance and the massive collinearity that it induces.

Fixed effect models of time series data are affectionately called "models of ignorance" because they estimate a set of parameters with no substantive meaning. In the present case with 361 school districts in the analysis (missing values means this is not a balanced pool), a fixed effects model will estimate 360 additional parameters, one for each district (minus the omitted category). Before implementing a fixed effects model, however, the analyst should examine how much variance such a model soaks up and whether sufficient residual variation remains to provide a fair test for any substantive hypotheses. When we estimated a fixed effects model with only time dummies, an intercept, and the fixed effects as independent variables and "all students" test scores as the dependent variable, we explained 92% of the variance. In short, Nielsen and Wolf counsel us to throw away 92% of the variation in student test scores and perform our substantive tests on only 8% of the total variation in the dependent variable. We think that using any model that assigns 92% of the variance in the dependent variable to a set of parameters of no substantive meaning is always a poor choice (see also Beck and Katz 1996).

What makes a fixed effects model in this case the wrong choice rather than just a poor one is collinearity.[1] Nielsen and Wolf dismiss this concern and report variance inflation factors (VIF) for a few variables in their footnote 6. Their collinearity diagnostics, however, are incorrect. Collinearity is assessed by regressing an independent variable on all other independent variables; in a fixed effects regression these other independent variables include the dummy variables (Greene 1997, 421). One minus the R^2 from this equation is the tolerance, and the VIF is the reciprocal of the tolerance (Maddala 1992, 274).[2] Table 1 presents VIFs for our original OLS model, the model with percent Anglo students added, and the fixed effects model, along with the VIFs that Nielsen and Wolf report for this same fixed effects model. While collinearity is a concern in the first two models, in the fixed effects model it is debilitating. A rule of thumb for major collinearity problems is a VIF of 10 or more (Hamilton 1992, 135). Every variable in the fixed effects model except gifted classes exceeds this criterion; for percent Anglo students (232.56) and percent minority teachers (76.33) the VIFs are massive. Viewed another way, the R^2 between percent minority students and all other independent variables is .9869, for Anglo students .9957. With such massive collinearity, models must use the unique variation to generate parameter estimates (Davidson and MacKinnon 1993, 50-1; Fox 1997, 338). In this case the minority teachers variable has only 1.31% of its total variation available, but even that is substantial compared to the 0.43% for Anglo students. Nielsen and Wolf's VIFs not only are incorrect by orders of magnitude, they are smaller than the VIFs for the base OLS regression; by definition that is not possible; this alone should have led them to question the accuracy of their own figures.[3]

[Table 1 About here]

Massive collinearity in the fixed effects model, therefore, explains why all of the variables, variables common place in education production functions, become insignificant, flip signs, or are reduced drastically in magnitude (compare for example columns 1 and 8 in Neilsen and Wolf, table 1; see Fox 1997, 339; Hamilton 192, 134-5). In direct contradiction to their claim about the lack of insignificant coefficients, the fifteen significant coefficients in columns 1 and 2 of their table 1 compare unfavorably to the five in the fixed effects models (columns 7 and 8). Anyone examining these results with collinearity diagnostics in hand can place precious little confidence in the fixed effects models reported by Nielsen and Wolf.

The zen of panel model selection should also be a concern. How does one decide that a fixed effects model is appropriate? One includes the fixed effects in the regression and then does a joint significance test of the set of dummy variables (Greene 1997, 617-8). Nielsen and Wolf (N.d., n5) report a significant test with 349 dummy variables but do not report the full information from the test.[4] The zen of this test is that one uses solely statistical criterion to add variables of no substantive meaning to a regression model. <u>This specification test, therefore, is the panel analysis equivalent of stepwise regression</u>, an atheoretical technique that relies on computers rather than theory to build models.

Fixed effects models may have a use in the analysis of some policies, but the present case is not one of them. Nielsen and Wolf's fixed effects models are flawed by excessive collinearity induced by a set of 360 theoretically meaningless dummy variables. We conclude that the fixed effects approach undertaken by Nielsen and Wolf is inappropriate and their statistical results and substantive inferences from these models lack veracity. Results based on these models can be dismissed from further consideration.

2. The Exclusion of Percent Anglo Students

Nielsen and Wolf contend that we have underspecified our models by omitting a key variable–the percent of Anglo students. We need to put this argument in context. Originally Nielsen and Wolf (1999) claimed that we omitted important peer effects and argued that students who attend school with better students are improved in the process. While we agree with this contention, we reject their original notion that better students means Anglo students. We did not even consider using percent Anglo students in our analysis for several reasons. First, the focus of our analysis was the impact of minority teachers on minority students and Anglo students (as opposed to overall student performance). Nielsen and Wolf do not cite a single study supporting the inclusion of percent Anglo students into production functions for minority students or Anglo students. In constructing our models we had no theoretical reason for including the percentage of Anglo students into either the minority or the Anglo equation.

Second, we are unclear what percent Anglo students measures; it clearly does not measure peer effects since as columns 1 and 2 of Table 2 demonstrate, the slope for this variable is consistently negative (only in the fixed effects models does it turn positive but with a VIF of

232.56 that can be discounted). In other words, both Anglo and minority students do worse (not better) in school districts with more Anglo students. While the notion that Anglos are inherently less able students is interesting, we are skeptical of this conclusion for methodological reasons (collinearity) and the lack of theory involved.[5]

[Table 2 About Here]

Third, the costs of using percent Anglo students are not compensated by any real gains. Adding this variable to our original "all students" regression increases the R^2 from .7055 to .7098 or .0043. In the nonlinear specification, the increase is from .7251 to .7261 or .0010. Although this additional explanation might be statistically significant, it is substantively trivial.

Fourth, if peer effects are a concern, why not simply add the actual performance of Anglo students to the minority student equation (see Weiher, 2000)? Doing so controls for the quality of Anglo peers but also creates an inherent comparison of groups in the equation. This comparison means the results are defined to be zero sum between the two groups so we can no longer assess any distributional aspects (that is, run the same equation for Anglos); thus, we limit our analysis to minority students only.[6] Column 3 of table 2 demonstrates that even controlling for peer effects, minority teachers have a positive association with minority student performance. Peer effects have a substantial impact in the equation and swamp the impact of some of the resources variables (but not that of minority teachers).

In reality the inclusion of percent Anglo students by Nielsen and Wolf only mimics our own logic for the puzzle that we found – positive relationships for Anglos and minorities but a negative one for all students. We concluded that this was because the overall scores were unduly influenced by districts with few minorities and illustrated this with residual analysis, follow-up regressions, and SWLS analysis. Our intent was to investigate an interesting puzzle, a puzzle peripheral to our core argument that all other things being equal both minority and Anglo students have higher test scores in districts with more minority teachers. At best Nielsen and Wolf probe a tangent to our paper; at worst they commit a specification error by introducing an irrelevant variable.

3. The Absence of a Nonlinear Relationship

Nielsen and Wolf contend that the relationship between all students test scores and minority teachers is not nonlinear.[7] Their method of analysis is curious since their own regression equation in table 1 (model 4) shows a nonlinear relationship even with percent Anglo students in the model. Their argument is a series of graphs where a multidimensional relationship is reduced to two-dimensional plots. They even bias their argument a bit by adding an additional 218 observations that did not meet our criterion for multiracial school districts.[8]

Nielsen and Wolf's argument has three serious flaws. First, we suggest the relationship between minority teachers and all students test scores is nonlinear when controlling for state aid, instructional funding, teacher salaries, gifted classes, class size, low income students, and teacher experience (we are even willing to add percent Anglo students here); we do not contend that the simple relationship between these two variables with no controls is nonlinear. Nielsen and Wolf's graphs can only show the bivariate relationship.[9] Second, even Nielsen and Wolf's visual argument is biased by manipulating the y-axis. By compressing the scale of the Y axis, they have made the relationship appear flatter (more linear) than it actually is. Third, the classic test of a nonlinear relationship is whether or not both a linear and a squared term are statistically significant. Our argument meets this test with flying colors in our specification (t-scores 14.12

and 12.21 for linear and nonlinear) and theirs (t = 11.39 and 11.11). Even with their specification (i.e., with percent Anglo students), the nonlinear relationships for all students, Anglo students, and minority students are all statistically significant (see Table 3).

[Table 3 About Here]

Nielsen and Wolf state "It is hard to envision a mechanism by which concentrating minority teachers in public school districts would suddenly make them more effective . . . " This contention reveals that Nielsen and Wolf are unfamiliar with the literature on representative bureaucracy which provides both theoretical and empirical evidence why a critical mass is needed for effective representation in some organizations. Theoretically, organizational incentives create pressures on minorities to conform and not press minority issues until their number is sufficient to provide support for such decisions (Thompson 1976; Meier 1993). The empirical literature provides several supporting findings: a large number of minority teachers is positively associated with the election of minority school board members (Meier and Smith 1994) and minority board members provide political support for minority teachers (Meier 1993; Henderson 1979), minority teachers are the recruitment base for minority administrators and a critical mass of administrators generates more active representation (Meier and Stewart 1991; Meier 1993), organizational processes in education, particularly grouping and tracking, are slow to change and changes in these policies is far easier when minorities become a majority (Meier and Stewart 1991; Meier, Stewart and England 1989), representation is a function of organizational roles (Selden, Brudney, and Kellough, 1997) and active representation roles are easier to establish in bureaucracies with more minorities.

4. Minorities are Inappropriately Grouped

Our analysis grouped blacks and Latinos; we did so because prior research demonstrated the political tradeoffs in multiracial situations (see Meier and Stewart 1991). Multiracial jurisdictions can create dynamics whereby Anglos initiate coalitions with one minority group at the expense of the other. Because our interest was Anglo-minority tradeoffs, we examined minorities as a group. At times simplification is useful to address interesting issues.

Even though this is the only criticism that Nielsen and Wolf lodge against our core findings, their arguments are based on the same flawed fixed effects statistical modeling approach that we demonstrated contains little empirical utility. Table 4 disaggregates the analysis by Latino and black students using our initial model. In both cases, the minority teachers in question are associated with positive results with their respective student groups. As a further test, we include Nielsen and Wolf's percent Anglo student variable and also the peer effects measure.[10] Again even with these controls, the relationship between black teachers and black students is positive and the relationship between Latino teachers and Latino students is positive.

[Table 4 About here]

5. The Level of Analysis

Nielsen and Wolf criticize our use of school districts as the unit of analysis expressing a clear preference for analysis at the individual student level. On one level this is a red herring since Nielsen and Wolf know the full set of these data does not exist at the individual level and even at the campus level key variables such as instructional funding per student do not exist.[11] All of their analysis is also at the district level.

At a second level the criticism is misinformed because it does not take into consideration the characteristics of the data generating process. Meier et al. (1999: 1030) explicitly state three ways that minority teachers might affect minority test scores:

> (1) minority teachers are more effective at teaching minority students (Aaron and Powell 1982, 5; Moore and Johnson 1983, 472); (2) minority teachers serve as role models for minority students (Cole 1986, 332); and (3) minority teachers mitigate the negative consequences of grouping, tracking, and discipline (Meier and Stewart 1991).

Why will an individual level analysis miss or underestimate each of these processes? First, examining the impact at the individual level essentially tests for the influence of a given teacher on a specific student in the year the test is taken. If minority student A has minority teacher B in the second grade and benefits greatly from the experience and this learning advantage persists, would it not also show up on the third grade exam or even the seventh or eighth grade exam? Perhaps this type of impact can be estimated with some relatively complex distributive lag models, but even the best individual level data sets in Texas or elsewhere do not permit researchers to link individual students unambiguously to individual teachers over a multiyear period. Second, if the process works via role models, a minority student need not have a class with a minority teacher to benefit. Minority teachers in other classrooms and even in other district schools are highly visible, potential role models, to students. The probability of this type of exposure can only be measured at the organizational level. Third, if the process works through decisions about grouping and tracking, the result is both likely to persist through time and is likely to be influenced by other "representatives" than that year's classroom teacher.

Our analysis was concerned with representation. Representation is a process that can occur either on the dyadic level (i.e., your Member of Congress acts in your interest) or at the collective level (your Member of Congress does not act in your interest, but other Members do even though they are not your representative, see Weissberg 1978). Collective representation is a concept that can only be examined at a collective level. Given these three processes of representation, quite clearly representative bureaucracy as applied to education means that it must be examined at the organizational level. Not only are organizations the appropriate level of analysis; but an individual level analysis, even if it were possible, would be unlikely to measure the representation process accurately.

6. The Incorrect Use of SWLS

We used SWLS, a form of substantively weighted analytical techniques (SWAT), to examine a subset of school districts to determine how the process of representation might work in districts with relatively equitable test scores. Nielsen and Wolf criticize our use of SWLS on two grounds (1) we should have selected a different subset of districts to examine and (2) we should have examined equity by using equity as a dependent variable in a traditional regression (Nielsen and Wolf, N.d., n5). Both arguments stem from a misunderstanding of the philosophy behind SWAT (see Meier and Gill 2000)[12] and the latter argument fails to consider the ways equity might be achieved.

SWLS is a qualitative technique that assumes the relationships among variables can vary across organizations or other units of analysis. For a variety of reasons, school A might be able to produce better results with $4000 per student than school B. SWLS asks the question, if there were a different set of organizations with a given set of characteristics, how might the overall relationships change? We examined those districts with relatively equitable test scores (minority vs. Anglo) to determine how minority teachers might influence test scores in this type of districts. The key finding is presented as a graph on page 1036 in the original article. This graph shows the slopes for minority teachers increasing for both Anglos and minorities and decreasing for all students as the equity weights increase. Since the all students slope is negative, this finding corroborates other results in regard to the all students relationship.

Examining the least equitable districts, Nielsen and Wolf contend that minority teachers are less important, perhaps even unimportant, and that other factors increase in importance. This is hardly a criticism of SWLS or our use of it. SWLS was designed to make such assessments. Meier and Gill (2000, 59-78) illustrate exactly this process. Nielsen and Wolf's misunderstanding of SWLS means that our point of contention is whether it is better to focus on equitable or inequitable districts. In reality we think it best to examine both, but examining districts that get good results while maintaining interracial equity (our set) seems at least as defensible as districts that get good results by exacerbating the gap between minority and Anglo students.

Nielsen and Wolf's second criticism is that we should directly examine equity and in that case equity would be negatively related to minority teachers. Their criticism overlooks the obvious–a district can achieve equity in two ways by improving minority scores until they equal Anglo scores or by decreasing Anglo scores until they match minority scores.[13] Many equitable school districts are bad districts; they fail to educate students regardless of race. Other equitable districts are good districts that get high scores for both minorities and Anglos. Rather than group these unlike sets together, does it not make more sense to examine the ability to produce both high scores and equity, in short, to examine the ability to attain two goals simultaneously? The promise of SWAT is that it can provide information on just such questions. SWAT uses a different approach to analysis, a zen that requires the analyst to explicitly specify values and alternative scenarios and then to see how those scenarios might evolve from the data. Yes, when one uses a flexible technique such as SWAT, it can be misapplied; but as Nielsen and Wolf demonstrate, so can rigid techniques like Hausman specification tests. There are no right and wrong subsets to apply SWAT, only subsets that are more interesting than others. Our subset, districts that are equitable and perform well, is interesting; Nielsen and Wolf's subset, districts that perform well and are inequitable, is interesting as well. This discussion, however, should not detract from the main point. Our findings for both minority and Anglo students are based on traditional regression techniques. We use SWAT to probe an interesting aspect of those findings, but the findings themselves do not rely on the SWAT results.

Conclusion

This response demonstrates that Nielsen and Wolf fail to challenge our core findings: Representative bureaucracies can benefit both minorities and non-minorities alike. The bulk of Nielsen and Wolf's comments apply to tangential issues rather than these core findings. Even on these tangential issues, however, they propose well intentioned, yet seriously misguided estimation strategies, present erroneous results, and provide weak arguments for their empirical modeling that are grounded in neither theory nor substance.

This exchange illustrates the benefits of replication. Two different approaches to the same question have illustrated the logic and choices that scholars must make. This would make an excellent teaching example. Our data and complete documentation of the analysis are available on line.

Table 1. Collinearity Diagnostics: VIF Estimates

Independent Variable	Base Model	NW Model	NW FE	NW Reported
Minority Teachers	1.88	2.61	76.34	1.09
% Anglo Students	NA	4.45	232.56	1.90
Instruction Funds K	4.14	4.21	11.06	NR
Teacher Salaries K	4.19	4.30	19.92	NR
State Aid	1.57	1.58	18.42	NR
Gifted Classes	1.04	1.04	3.93	NR
Class Size	2.85	2.86	13.25	NR
Percent Low Income	2.29	3.76	22.12	1.75
Teacher Experience	1.25	1.33	10.06	NR

Base Model = Nielsen and Wolf, Table 1, Model 1

NW Model = Nielsen and Wolf, Table 1, Model 2

NW FE = Actual results for Nielsen and Wolf, Table 1, Model 8

NW Reported = Results reported by Nielsen and Wolf, Table 1, Model 8 note 6.

NA = not applicable

NR = not reported

Table 2. Anglo Students verses Peer Effects

Independent Variable	Nielsen-Wolf Anglo Students	Nielsen-Wolf Minority Students	Minority Students
Minority Teachers	-.0959 (.0190)	.0238ns (.0224)	.0876 (.0186)
% Anglo Students	-.2497 (.0149)	-.0554 (.0175)	.1106 (.0154)
Anglo Test Scores	----	----	.6652 (.0212)
Controls			
Instruction Funds K	-.0101ns (.8199)	2.8414 (.9632)	2.8481 (.7941)
Teacher Salaries K	.3499 (.1336)	1.3229 (.1569)	1.0902 (.1296)
State Aid	.0060ns (.0093)	-.4453 (.2031)	.0620 (.0090)
Gifted Classes	.2586 (.0484)	.1766 (.0569)	.0046ns (.0472)
Class Size	-.6800 (.1729)	-.4453 (.2031)	.0070ns (.1681)
Percent Low Income	-.3566 (.0183)	-.3009 (.0215)	-.0638 (.0193)
Teacher Experience	.8236 (.1116)	-.6567 (.1311)	-1.2045 (.1095)
R-Square	.60	.54	.69
Standard Error	6.63	7.79	6.42
F	220.84	178.03	309.91
N	2097	2097	2097

standard errors are in parentheses
annual dummy variables are not reported

Table 3. Nonlinearity and the Nielsen-Wolf Specification

Independent Variable	All Students	Anglo Students	Minority Students
Minority Teachers	-.4750	-.4550	-.4990
	(.0417)	(.0434)	(.0505)
Minority Teachers Squared	.0076	.0065	.0095
	(.0007)	(.0007)	(.0008)
% Anglo Students	.0386	-.2858	-.1080
	(.0145)	(.0151)	(.0176)
Controls			
Instruction Funds K	1.3994ns	.3790ns	3.4079
	(.7733)	(.8051)	(.9356)
Teacher Salaries K	.7573	.2470ns	1.1731
	(.1263)	(.1315)	(.1528)
State Aid	.0153ns	.0019ns	.0601
	(.0088)	(.0091)	(.0106)
Gifted Classes	.2543	.2327	.1389
	(.0457)	(.0476)	(.0553)
Class Size	-.5859	-.5873	-.3104ns
	(.1631)	(.1698)	(.1973)
Percent Low Income	-.3555	-.3638	-.3115
	(.0173)	(.0180)	(.0209)
Teacher Experience	.3352	.9221	-.5134
	(.1056)	(.1099)	(.1278)
R-Square	.73	.61	.57
Standard Error	6.24	6.50	7.55
F	367.72	219.92	185.38
N	2097	2097	2097

standard errors are in parentheses
annual dummy variables are not reported

Table 4. The Impact of Black and Latino Teachers on Similar Students

Independent Variable	Black Students		Latino Students	
Black Teachers	.1706 (.0453)	.2794 (.0464)	NA	NA
Latino Teachers	NA	NA	.1400 (.0280)	.0996 (.0291)
% Anglo Students	----	.0602 (.0284)	----	.1326 (.0232)
Anglo Test Scores	----	.4657 (.0462)	----	.5827 (.0352)
Controls				
Instruction Funds K	.2113ns (1.7051)	-.2674ns (1.6776)	2.6953ns (1.3862)	2.6769 (1.3157)
Teacher Salaries K	1.5547 (.2649)	1.3617 (.2711)	.8434 (.2278)	.8427 (.2171)
State Aid	.0779 (.0195)	.0820 (.0191)	.0324 (.0160)	.0344 (.0151)
Gifted Classes	-.2181 (.1014)	-.3351 (.0997)	.1727 (.0827)	.0412ns (.0782)
Class Size	.9132 (.3587)	1.1051 (.3543)	-.1587ns (.2911)	.3497ns (.2772)
Percent Low Income	-.3202 (.0241)	-.2062 (.0410)	-.3004 (.0227)	-.0591ns (.0321)
Teacher Experience	.2979ns (.2288)	-.0841ns (.2392)	-.2881ns (.1890)	-.8498 (.1835)
R-Square	.29	.32	.39	.46
Standard Error	13.91	13.58	11.31	10.63
F	65.68	66.74	102.05	118.46
N	2097	2097	2097	2097

standard errors are in parentheses
annual dummy variables are not reported

References

Aaron, Robert and Glen Powell. 1982. "Feedback Practices as a Function of Teacher and Pupil Race during Reading Group Instruction." *Journal of Negro Education* 51 (Winter): 50-59.

Baltagi, Badi H. 1995. *Econometric Analysis of Panel Data*. New York: John Wiley and Sons.

Baltagi, Badi H. 1999. *Econometrics*. 2nd Revised Edition. New York: Springer.

Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6: 1-36.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series-Cross-Section Data." *American Political Science Review* 89 (September): 634-647.

Cole, Beverly P. 1986. "The Black Educator: An Endangered Species." *Journal of Negro Education* 55 (Summer): 326-334.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.

Greene, William H. *Econometric Analysis*. 3rd. Englewood Cliffs: Prentice-Hall.

Hamilton, Lawrence C. 1992. *Regression with Graphics*. Belmont, CA: Brooks/Cole.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Maddala, G. S. 1992. *Introduction to Econometrics*. 2nd Edition. New York: Macmillan Publishing Company.

Meier, Kenneth J. 1993. "Latinos and Representative Bureaucracy: Testing the Thompson and Henderson Hypotheses." *Journal of Public Administration Research and Theory* 3 (October): 393-415.

Meier, Kenneth J. and Jeff Gill. 2000. *What Works: A New Approach to Program and Policy Analysis*. Boulder, CO: Westview Press.

Meier, Kenneth J., Robert D. Wrinkle, and J.L. Polinard. 1999. “Representative Bureaucracy and Distributional Equity: Addressing the Hard Question.” *Journal of Politics* 61 (November):1025-39.

Meier, Kenneth J. and Kevin B. Smith. 1994. “Representative Democracy and Representative Bureaucracy.” *Social Science Quarterly* 75 (December): 798-803.

Meier, Kenneth J. and Joseph Stewart, Jr. 1991. *The Politics of Hispanic Education*. Albany: SUNY Press.

Meier, Kenneth J. Joseph Stewart, Jr. and Robert E. England. 1989. *Race, Class and Education*. Madison: University of Wisconsin Press.

Moore, Helen A. and David R. Johnson. 1983. "A Reexamination of Elementary School Teachers' Expectations: Evidence of Sex and Ethnic Segmentation." *Social Science Quarterly* 64 (September): 460-475.

Nielsen, Laura and Patrick Wolf. N.d. “Representative Bureaucracy and Harder Questions: A Response to Meier, Wrinkle, and Polinard.” *Journal of Politics* (forthcoming).

Nielsen, Laura and Patrick Wolf. 1999. “Representative Bureaucracy and Harder Questions: A Response to Meier, Wrinkle, and Polinard.” typescript, Georgetown University.

Selden, Sally Coleman, Jeffrey L. Brudney, and J. Edward Kellough. 1998. "Bureaucracy as a Representative Institution." *American Journal of Political Science* 42 (July): 717-44.

Thompson, Frank J. 1976. "Minority Groups in Public Bureaucracies." *Administration and Society* 8 (August) :201-226.

Weiher, Gregory. 2000. “Minority Student Achievement: Passive Representation and Social Context in Schools.” *Journal of Politics* 62 (August), 886-895.

Weissberg, Robert. 1978. "Collective vs. Dyadic Representation in Congress." *American Political Science Review* 72 (June), 535-547.

Biographies:

Kenneth J. Meier is Charles Puryear Professor of Liberal Arts and Professor of Political Science at Texas A&M University.

Warren S. Eller is a PhD candidate in political science at Texas A&M University.

Robert D. Wrinkle is Professor of Political Science at the University of Texas Pan American.

J.L. Polinard is Professor of Political Science at the University of Texas Pan American.

1.Baltagi (1999, 309) specifically warns about collinearity problems in fixed effects models using a 50 x 10 pool (one less likely to be problematic than the current one) as an illustration.

2.Analysts should perform these calculations rather than relying on computer software packages because some packages use computational tricks rather than working with the full data set. The problems with pooled software in particular have been demonstrated by Beck and Katz (1995).

3.The simple correlation between minority teachers and Anglo students (-.76, see Nielsen and Wolf, 1999) alone generates a larger VIF (2.36) than the one they report. This is an obvious indicator that their VIFs were calculated incorrectly because by definition the VIFs for minority teachers and Anglo students must equal or exceed 2.36 in any regressions where both appear.

4.We think Nielsen and Wolf confuse the Hausman test between fixed and random effects (see Greene 1997, 633) with the joint significance test which compares fixed effects to OLS (Greene 1997, 617). We also have no idea why they do a test for 350 school districts. This study has 361 school districts and district identification codes are clearly designated in the data set. Any test with 350 would not be a true specification test because it omits 11 cross sectional units. The original article misreported the number of districts as 350 but this should not affect the results of someone working from the data set. Our OLS vs. fixed effects test produced an F (360, 1722) value of 12.325; the Hausman test was also significant (chi square = 176.88, 9 df). Our web documentation file includes these tests and our estimate of the fixed effects model. Our estimates in STATA (verified in limdep) do not correspond with Nielsen and Wolf's. That file contains the data and documentation so that others can verify our results for themselves.

5.On the tangential issue of including this measure for the "all students" regression, Nielsen and Wolf would have an argument except for the overlap between Anglo students and minority teachers. Minority teachers are concentrated in school districts with more minority administrators and a larger minority population (Meier and Stewart 1991; Meier, Stewart and England 1989). Both factors are negatively correlated with Anglo students.

6.This comparative specification means that the estimate is likely an underestimate if minority teachers also benefit Anglo students.

7.Our argument does not rely on nonlinear relationships. This is merely an interesting tangent.

8.Why they added 218 cases is unclear since an additional 303 cases actually meet their criterion.

9.This graphical under specification is the same sin they lodge against our regression models.

10.We cannot think of a reason to have both Latino and black teachers in the same equation unless one wants to examine intraminority tradeoffs.

11.Weiher (2000) using different models replicates our findings for minority students at both the district and the campus level.

12.Meier and Gill (2000) was published in April 2000 three months before the revised version of Nielson and Wolf was submitted to this journal. The full text was available on the web in April 1999 with individual chapters available well before then. Numerous papers were presented at conferences or published in journals beginning in 1996. That Nielsen and Wolf cite none of these works and rely on a terse two page summary of SWAT in a more general article on a variety of methods is distressing. Their comment that we should have used regression-based residual weights misperceives the flexibility of SWAT. Meier and Gill (2000, 99-115) explicitly discuss the use of endogenous versus exogenous (the case here) weights.

13.Nielsen and Wolf's set of districts are among the worst in Texas. Our districts have an Anglo pass rate of 71.6 and a minority pass rate of 56.6; their districts' rates are 60.3 and 23.3 respectively. Their districts are both bad and inequitable.