DOCUMENT RESUME

ED 392 837 TM 024 608

| | |
|---|---|
| AUTHOR | Kreft, Ita G. G. |
| TITLE | The Effects of Centering in Multilevel Analysis: Is the Public School the Loser or the Winner? |
| PUB DATE | Apr 95 |
| NOTE | 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995). |
| PUB TYPE | Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; Elementary Secondary Education; *Estimation (Mathematics); Goodness of Fit; Mathematical Models; *Private Schools; *Public Schools; Research Methodology; *School Effectiveness |
| IDENTIFIERS | *Centering; *Multilevel Analysis; National Education Longitudinal Study 1988 |

ABSTRACT

Centering is a problem in multilevel analysis that needs to be addressed. Multilevel analysis gives researchers a choice between centering within context and grand mean centering, both statistically sound ways to improve estimation of the parameters in the model. Researchers of the most widely used software package for the analysis of hierarchically nested data, HLM, commonly center within context. The practice of centering around the group mean, while not adding the mean back into the model deletes information from the data, may lead to overestimation of the macro level variables in the model. Centering around the group mean is fitting another model. Reasons to center or not to center are discussed, based on examples from data from the National Education Longitudinal Study of 1988. Comparison of data from public and private schools about mathematics achievement shows that different conclusions are reached depending on the data management scheme. Although research on school effects is not yet free of methodological and conceptual problems, existing models have the potential to expand knowledge when used correctly. (Contains 6 tables and 17 references.) (SLD)

# The Effects of Centering in Multilevel Analysis: Is the Public school the loser or the winner?

A new analysis of an old question

Ita G. G. Kreft
California State University



Presented at AERA 1995, Section D
San Francisco, April 18-22 1995
Session Title: Multivariate Analysis of NELS:88

Correspondence should be addresses to
Ita G.G. Kreft
School of Education
California State University
5151 State University drive
Los Angeles, CA 90032

Fax 213 - 343 4318
Phone: 213 - 343 5116
email: kreft@math.ucla.edu

# ABSTRACT

Multilevel models raise new problems that need solutions, one of them is centering. Centering needs to be addressed for several reasons. Firstly multilevel models give researchers two centering choices, centering within context, and grand mean centering, both statistically sound ways to improve estimation of the parameters in the model. Secondly, researchers of the most widely used software package for the analysis of hierarchically nested data, HLM, commonly center within context. The practice of centering around the group mean, while not adding the mean back into the model deletes information from the data, which may lead to overestimation of macro level variables in the model. Centering around the group mean is fitting another model. Centered predictors are not equivalent in meaning or interpretation to raw score predictors, and analyses will yield different results. Applying centering within context or using raw scores serves a purpose, and needs a scenario. Based on that scenario reasons for to center or not to center will be given. The examples in this paper are based on the NELS88 data, where math achievement is predicted by Homework, SES and Sector, either public or private, the same variables as used in Raudenbush and Bryk (1986), producing the same result if centering is used. This is no longer true when raw scores are used. The conclusion reached, is that multilevel analysis cannot determine if the private sector is better than the public sector, since it all depends on how the data is treated. Based on analysis with the same data, but applying different data management schemes, opposite conclusions are reached in relation to the effect of Sector. The final conclusion is, that the stage has not been reached, where research on school effects is "no longer plagued by methodological and conceptual problems", as Raudenbush and Bryk in 1986 hoped for. But the models have the potency to expand knowledge in the field, when used in the proper way.

# INTRODUCTION

## The scaling of micro level predictors

Random Coefficient models are models for analyzing grouped data. The data are assumed to exist on two or more levels, with the lower level nested within the higher level(s). Examples include students nested within schools, employees nested within firms, or alternatively, repeated measurements nested within persons. The lowest level measurements are said to be at the micro level; all higher level measurements, at the macro level. Macro levels are often referred to as contexts. Models may have as few as two levels, as in the case of students (micro level) nested within classes (macro level); or more than two, for example, a three-level model of students nested within classes, nested within schools. Measurements at the micro level are assumed to vary both within each context and among contexts.

Some controversy exists as to how the micro-level variables should be scaled for multilevel analyses (Kreft, De Leeuw and Aiken, 1995). The available software packages for the analysis of hierarchically nested data, GENMOD, HLM, ML3 and VARCL (see Kreft, De Leeuw and Van der Leeden, 1994, for an overview). They differ in the way they handle the raw data. The option in HLM (Bryk, Raudenbush, Selzer and Congdon, 1988), and preferred by the authors of the software, (see e.g. Bryk and Raudenbush (1992, p.25 etc.) is to center predictors around the context mean, when coefficients for these predictors are assumed to be random. ML3 (Rasbash, Prosser and Goldstein, 1989) offers choices with no outspoken preference. VARCL centers around the grand mean, but put the results back into raw score solutions. The manual for ML3 specifies reasons for centering, as it may facilitate interpretation, but mainly as useful to improve numerical performance of the estimation algorithm. No software manual, nor other publications report the difference in the macro parameter estimates obtained by centering around the context mean, as compared to raw score solutions. The only publication known to me is Kreft, DeLeeuw and Aiken (1995).

Micro-level measures may be left in raw score (RS) or may be centered, i.e. converted to deviations from the mean (CWC) or centered around the grand mean of the full data set. Centering in any of these two forms improve computational ease (Longford, 1990) and improve estimation by reducing multi-collinearity. But more importantly for the user, the scaling of micro-predictors around the group mean has consequences for the parameters estimates in the model. Deviation scores, instead of raw scores, used by HLM practitioners, are subsequently discussed in their analyses, as if the original raw scores were used (e.g. Lee and Bryk, 1988, and Raudenbush and Willms, 1991). An inattentive reader may not realize that the variables in these models are no longer raw scores Since centering changes relationships in the model, especially the relationship between the dependent variable and macro level variables, the practice of centering and pretending that raw scores are used is misleading. In this paper I explore the impact of centering within context in multilevel analysis and its different effects on conclusions reached regarding the private and public sector. For a better understanding the equations of random coefficient (RC) models are introduced first.

## The random coefficient model

In micro-equation (1), which is familiar in the literature on hierarchical linear modeling (Kreft, DeLeeuw and Kim, 1990):

$$(1) \qquad \underline{Y}_{ij} = \underline{a}_j + \underline{b}_j X_{ij} + \underline{e}_{ij}, \text{ where } \underline{e}_{ij} = N(0, \sigma^2)$$

The coefficients in these models are modeled with an error term which represents the uncertainty inherent in observing only a sample of all possible contexts. For that reason we use the term random coefficient (RC) models, instead of hierarchical linear models (HLM) in order to avoid the common mistake that we talk about the software package HLM. In equation (1) and all equations that follow we adopt the convention of underlining random coefficients. Index i is used for individuals, index j is used for contexts. Variable $\underline{a}_j$ is the random intercept, $\underline{b}_j$ is the random slope, and $\underline{e}_{ij}$ is the micro level error term (disturbance). We assume that $\underline{e}_{ij}$'s have expectation zero, are independent, and are normally and identically distributed. The variance of $\underline{e}_{ij}$ is equal to $\sigma^2$. In model (1) there is no measure of the macro-level units beyond the identity of which individuals belong to which contexts. Hence the macro-level equations express the properties of the random slope and intercept in terms of overall population values plus error, as specified in the macro equations (2) and (3):

$$\underline{a}_j = \gamma_{00} + \underline{\delta}_{0j} \quad (2)$$

$$\underline{b}_j = \gamma_{10} + \underline{\delta}_{1j} \quad (3)$$

The macro-level errors (disturbances) $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ in (2) and (3) respectively, indicate that both the intercept $\underline{a}_j$ and $\underline{b}_j$ vary over contexts. The grand mean effect in (2) is $\gamma_{00}$, while $\underline{\delta}_{0j}$ (the macro-error term) measures the deviation of each context from this overall or grand mean. The same is true in (3) where the grand slope estimate across all contexts is $\gamma_{10}$, while $\underline{\delta}_{1j}$ represents the deviation of the slope within each context from the overall slope. For the gamma's the subscript is defined as follows: the first index is the number of the variable at the micro level, the second represents the number of the variable at the macro level. Hence $\gamma_{st}$ is the effect of the macro level t on the regression coefficient of micro variable s. Zero signifies the intercept, i.e. the variable with all values equal to +1, either at the micro level or at the macro level. For instance $\gamma_{00}$ is the effect of the macro level intercept on the micro level coefficient of the intercept. Note that (2) and (3) display the model coefficients $a_j$ and $b_j$ as a function of two components: a fixed component $\gamma_{00}$ and $\gamma_{10}$ respectively, and a random component $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ respectively, where $\underline{\delta}_{0j}$ has variance $\omega_{00}$, $\underline{\delta}_{1j}$ has variance $\omega_{11}$, while $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ have covariance $\omega_{01}$.

In one of our analyses of math achievement predicted by hours of homework, the micro-equation corresponding to (1) above is as follows:

math achievement = $\text{intercept}_j + b_j$ homework + error

Since the intercept and the b-coefficient are allowed to differ over schools, an overall intercept and slope ($\gamma_{00} + \gamma_{10}$ in equations 2 and 3) are estimated, from which each school is allowed to differ.

This dispersion is shown in the error or random part ($\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$) of the equations (2) and (3) for intercept and slope, respectively.

The variances of the macro–errors $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ and their covariance are parameters of the model and are given in the matrix $\Omega$. The terms in $\Omega$ are referred to as <u>variance components</u> of the model. For the omega's the subscripts all refer to macro level variables. This means that $\omega_{st}$ is the covariance between random regression coefficients s and t. Zero refers again to the random intercept.

$$
(4) \qquad \Omega = \begin{array}{c|cc} & \underline{\delta}_{0j} & \underline{\delta}_{1j} \\ \hline \underline{\delta}_{0j} & \omega_{00} & \omega_{01} \\ \underline{\delta}_{1j} & \omega_{10} & \omega_{11} \end{array}
$$

The random fluctuation of the industry slopes around the mean slope is measured by variance $\omega_{11}$; fluctuation in intercepts over industries is measured by variance $\omega_{00}$, while the covariance between $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ is $\omega_{10}$.

Substituting the separate equations (2) and (3) into (1) produces the single equation (5a) of the RC model.

$$(5a) \qquad \underline{Y}_{ij} = (\gamma_{00} + \underline{\delta}_{0j}) + (\gamma_{10} + \underline{\delta}_{1j})X_{ij} + \underline{e}_{ij}$$

Expanding and rearranging terms in (5a) yields (5b).

$$(5b) \qquad \underline{Y}_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + (\underline{\delta}_{0j} + \underline{\delta}_{1j}X_{ij} + \underline{e}_{ij})$$

Equation (5b) resembles a fixed effect linear model with a complicated error term (within parenthesis).

If we introduce a macro–level predictor $Z_j$ into the intercept equation, yielding (5c) below, while retaining micro-equation (1) as before and macro-equation (3) for the slope as before: now rewritten as (5d):

$$(1) \qquad \underline{Y}_{ij} = \underline{a}_j + \underline{b}_j X_{ij} + \underline{e}_{ij} \qquad \text{where}$$

$$(5c) \qquad \underline{a}_j = \gamma_{00} + \gamma_{01}Z_j + \underline{\delta}_{0j} \qquad \text{and}$$

$$(5d) \qquad \underline{b}_j = \gamma_{10} + \underline{\delta}_{1j}$$

In (5c) the intercept $\underline{a}_j$ of each context is now shown to be a function of both the group level variable $Z_j$ and random fluctuation $\underline{\delta}_{0j}$. The slope is as before a random coefficient without a relation to a context variable. Substituting (5c) and (5d) into (1) yields (5e), a combined equation analogous to (5a):

$$(5e) \qquad \underline{Y}_{ij} = \gamma_{00} + \gamma_{01}Z_j + \underline{\delta}_{0j} + (\gamma_{10} + \underline{\delta}_{1j})X_{ij} + \underline{e}_{ij}$$

Expanding and rearranging terms in (5e) yields (5f) analogous to (5b) above:

$$(5f) \qquad \underline{Y}_{ij} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{10}X_{ij} + (\underline{\delta}_{0j} + \underline{\delta}_{1j}X_{ij} + \underline{e}_{ij})$$

Once again, this appears like a fixed effects regression equation with a complex error term.

The covariance matrix $\Omega$ of the macro disturbances is as before. Also as before, the model parameters include the <u>regression</u> <u>estimates</u>

$\gamma_{00}$, $\gamma_{10}$, and $\gamma_{11}$; and the <u>variance components</u>, $\sigma^2$ at the micro level, and the elements of $\Omega$ at the macro level.
It should be noted that the combination of the CWC micro-predictor $(X_{ij} - X_{.j})$ with the macro-predictor $(X_{.j} - X_{..})$ produces the familiar partitioning of total variance in micro predictor $X_{ij}$ into a within context component and a between context component, i.e.:

$$X_{ij} = (X_{ij} - X_{.j}) + (X_{.j} - X_{..}) + X_{..}$$

The macro-predictor can be added to (5d) for the prediction of the random slopes $\underline{b}_j$., which leads to a model containing an explicit interaction between X and Z. Such a model is used by Raudenbush and Bryk, where the macro level variable Sector (Z) interacts with the micro variable SES (X), as in (5g) and (5h).

(5g) $$\underline{b}_j = \gamma_{10} + \gamma_{11} Z_j + \underline{\delta}_{1j}$$

Substituting (5g) and (5c) in (1) yields (5h), which differs from (5f) only in the interaction term $X_{ij}Z_j$.

(5h) $$\underline{Y}_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + \gamma_{11} X_{ij} Z_j + (\underline{\delta}_{0j} + \underline{\delta}_{1j} X_{ij} + \underline{e}_{ij})$$

The parameters of interest are the macro-level estimates of the $Z_j$'s in equation (5c) to (5h), and the variance components in equation (4), when comparing raw score models with models that center micro-level predictors around the group mean.

## CENTERING

The NELS88 data are used as illustration. The predictors are chosen based on the Raudenbush and Bryk (1986) analyses with the High School and Beyond data. At the student level they are: math achievement as the dependent variable and socioeconomic status (SES) and Homework as the student level predictors. Homework is "hours of homework per week" ranging from '0' to 'more than 10'. SES is a composite of four variables for father's and mothers's educational level and income. These two predictors are used in analyses as raw score or as deviation scores. The deviation scores are calculated by subtracting the mean of the school from the raw score of each student for Homework and SES. It is up to the researcher to add this mean back to the model or not. The acronym CWC (Centered Within Context) indicates that centered first level predictors are used in the model. If not centered a raw score model is used, indicated by RS. As second level predictors we used Public versus Private sector (Public=1, Private=0), and Student/Teacher ratio.

If means of centered predictor variables are reintroduced in the equation they are introduced as macro level predictors $Z_j$'s. The mean over a variable $X_{ij}$ is defined as

$$X_{.j} = 1/n_j \, \Sigma X_{ij}, \qquad \text{where } n_j \text{ is a number } i \in I_j$$

or $X_{.j}$ is the mean of observations within each school.

Centering within context yields different result for second level predictors compared to a raw score model. The habit of many HLM users, to center without introducing the school mean back in the model, leads to a substantial loss of important variance and to a worse model-fit. The reason for this loss is very straightforward: $X_{ij}$ as a raw score micro-predictor contains, among others, any variation in context means $X_{.j}$ over contexts, while this source of variation is removed from the micro-predictor ($X_{ij}$ - $X_{.j}$) in a centered analysis. If the variation among context means $X_{.j}$ is reintroduced at the macro-level, by using $X_{.j}$ as the macro-predictor, all variation in the original variable is present in the model, only divided up in a different way, and of course resulting in different solutions. A model with ($X_{ij}$ - $X_{.j}$) as micro-predictor and $X_{.j}$ as macro-predictor contains the same variation as a RS model with no context mean. The difference between CWC and RS model in that case is in the partitioning over the parameters only (see Kreft, DeLeeuw and Aiken, 1995). The same is not true for a CWC model, without the mean reintroduced.

## EXAMPLES

Throughout this paper we use the NELS88 dataset, with 21580 students in 1001 schools. The majority of these schools are public (80%) and the rest is divided over Catholic schools (10%) religious private schools (4%) and non religious private schools (6%). In the first analyses micro-level predictors, Homework and/or Socioeconomic Status, predict math achievement. In later analyses two macro level predictors Sector and Teacher/student Ratio are added with cross-level interactions.

Homework in deviation of the school mean, predicting math score.

In the first analyses Homework predicts math achievement. Two different analyses are compared. Homework in deviation of the school mean, with and without a reintroduction of the school mean in the model. It is not unusual to see applications of RC models (see for examples Willms and Raudenbush, 1992), where the mean is not added to the model, although the scores are put in deviation of the group mean. In Table 1 it shows that the difference in deviance between both models is large, indicating that model 2 where the subtracted mean is reintroduced, fits better (with a difference of 321 and 1 dfr). In general it is true, that leaving the mean out of a model with deviation scores is throwing away information that may be of importance. Later it is shown that omitting the mean will also change the results of the analysis for the macro coefficients.

Table 1. Homework predicting Math Achievement in CWC

1.CWC model without mean:

$$\text{Math} = 50.8\ (0.17) + 1.40\ (0.05)(\text{Homework}-X_{.j})$$

Variance intercept 26.81 (1.37). Variance slope: 0.60 (0.11). Deviance 155849

2. CWC model with mean:

$$\text{Math} = 40.85\ (0.53) + 1.40\ (0.05)(\text{Homework}-X_{.j}) + 5.08\ (0.26)(X_{.j})$$

Variance intercept 18.19 (0.98). Variance slope: 0.60 (0.11). Deviance 155528.

The difference between model parameter estimates in Table 1 is in the value for the intercept as well as its variance, which are considerably reduced by the introduction of the mean homework in the model. The magnitude for the coefficient of the micro predictor and its standard error is unchanged.

The Public school effect in raw score data compare to centered predictors

In Table 2 the effects of centering the micro predictor Homework is shown, on a macro level variable $Z_j$ ($Z_j$ is Sector, Coded as Public=1, Private=0). The models compared are the raw score model and two CWC models, one with and one without the mean reintroduced.

Table 2. Raw versus centered scores with macro level variable Public

| | | | |
|---|---|---|---|
| 1. RS: | Math =43.89 + 1.44**Homework | -2.95 (0.18)Public. | Deviance: 155478 |
| 2. CWC: | Math =54.99 + 1.38**(Homework-$X_{.j}$) | -5.33 (0.39)Public. | Deviance: 155590 |
| 3. CWC: | Math =45.07 + 1.39**(Homework-$X_{.j}$) | -3.61 (0.36)Public+ | |
| | + 4.38 (0.26) (Homework-Mean): | | Deviance: 155378 |

Note: the standard errors of the coefficient for Public are (between parenthesis)
The effect of Homework is still significant, as before. ** is significant at p=0.01

In all three models of Table 2 the Public sector has a negative effect on math achievement, after controlling for homework. The last two models in Table 2 show that centering strengthens that negative sector effect. The smallest effect for Public is in the raw score model (-2.95), while the largest effect (-5.33) is in Model 2, centering without reintroducing the mean. The effect of Public has different values over the tree models in Table 2, but all are of the same sign. The fit of the models differ too, as shown in the reported Deviances in Table 2. The model without the Mean reintroduced (model 2) has the worst fit, while the CWC model, with Mean (model 3), has the best fit. In the following analyses the negative Public sector effect disappears totally in a model where SES of the students is one of the predictors, and centering is applied.

Is the Private School a Winner, or a Looser?

In the analyses reported in Table 3 the variable SES is added to the model. This model resembles closely the model introduced by Raudenbush and Bryk (1986) for the Highschool and Beyond data, where math achievement was predicted by SES, Homework and Private/public Sector.

Table 3. The Different Effects of Sector over Different Forms of Centering

1. RS:

Math=50.16 + 1.24**Homework + 4.35** SES - 2.06 (0.29) Public

Variance slope SES: 0.49 (0.33) N.S. Deviance: 153333

2. CWC, without Means:

Math=55.06+ 1.18**(Homework–$X_{.j}$) + 3.84** (SES–$X_{.j}$) - 5.42 (0.34) Public

Variance slope SES: 0.4 0 ( 0.09) Deviance: 153968

3. CWC, with Means:

Math=47.53 + 1.20**(Homework–$X_{.j}$) + 3.85** (SES–$X_{.j}$) + 0.62 (0.28) Public
+ 1.65 (0.20) Homework-Mean + 8.14 (0.25) SES-Mean

Variance slope SES: 0.4 0 ( 0.09) Deviance: 153004

Note: the standard errors of the coefficient for the macro level variable are (between parenthesis)
** is significant at p=0.01 NS=Not Significant
* is significant at p=0.05

Centering variables has an effect on the value of the micro coefficients, but not substantial. The Homework and SES coefficients stay comparably large and equally significant. But centering predictors has an effect on macro level coefficients such as the coefficient for Public. The Public effect is large in the CWC model where the means are omitted (-5.42 compared to -2.02 in the RS model). However, a more dramatic difference in Public school effect is observed when comparing CWC model (2) and CWC model (3). The sign has changed to the advantage of the public school, from - 5.42 to +0.62. Depending on what model is used, or if the means are added back to the model, the conclusion is either that the public sector has a very strong negative effect on math achievement, even after correction for homework and SES, or has a positive effect (model 3). Not reintroducing the means of Homework and SES in the model favors strongly the Private sector. The effect for Public becomes more than twice as large compared to the model that uses raw scores (model 1). The deviances for the models differ too, and again the best fit is the model with centered variables, with means reintroduced. Centering predictors does not always produce a better fitting model, a finding that cannot be generalizedentered. The choice between models in this example will be in favor of CWC if a best fit is the criterion for the choice. But macro variables, such as means, serve a purpose, or have to be explained with a theory. Both models, CWC and RS, are different models and serve different purposes.

Another effect of centering observed in the analyses of Table 3 is on the variance for the slope of SES. In RS the slope for SES is not significantly random (variance 0.49, with a large standard error of 0.33), while the slope for SES is significant in both centering models (models 2 and 3 with a variance of 0.40 with a small standard error of 0.09). Centering, as we know, has an effect on the covariance between slope and intercept, and that may have on its turn an effect on the shrinkage of the b-coefficients for separate schools. This effect needs more investigation. For the moment, following guidelines in the manual for the software HLM, I would say, the slope is not significantly random, and as a consequence that slope is not centered. In HLM it is advised to center only predictors with random coefficients (advocated for the computational ease and the lowering of the covariance slope/intercept). Based on this advice I conclude that the relation SES-Math Achievement is not different over schools, if that difference does not exists in the raw score model (see model 1). I think that the significant variation of the effect of SES on the dependent variable in different schools in the CWC models is more a function of the different school means, rather than an interesting phenomenon that can be related to macro level differences such as Public versus Private. If the mean differs significantly, like in this example, subtracting the mean from each score will introduce a significantly different relationship among schools between SES and Math Achievement. In any case, SES and SES in deviation ($SES-X_{.j}$) can no longer be considered as the same variable.

An analysis with Student-Teacher Ratio added

To see if the same happens with another macro-level variable, one more school level variable is added to the model, Teacher/student Ratio. The higher the ratio, the higher the number of students per teacher. I assume that smaller classes are beneficial for students, and that this variable will show a negative effect on math achievement.

As before, the intercept and the two slopes for SES and Homework are defined as random. Again we find the slope for SES not significantly random in the RS model, but significantly random in the CWC model (z=3.89). The slope for homework is not affected by centering, and stays in all situations significantly random. Note that the micro-level regression coefficients in the analyses resemble closely the ones of Table 3. As before, the coefficient for Public changes from large and negative to positive in the deviation score model with added means. In the RS model Public it is highly significant (in favor of the private school). Using the RS model, a researcher would conclude that being in the Private sector and in classes with low Teacher/Student Ratio, have a positive influence on math achievement. Using a CWC model without means reintroduced reinforces this conclusion, since both macro effects are at least twice as strong as before. However, using a CWC model with the means reintroduced show totally different conclusions. An opposite effect for the Public sector, which becomes positive instead of a negative, and a much lower effect for Teacher/Student ratio.

Table 4. Adding a second level variable: Teacher/Student Ratio

1. RS:

$$\text{Math}=53.13 + 1.24^{**}\text{Homework} + 4.34^{**}\text{SES} -2.38^{**}(0.35)\text{Public} - 0.15^{**}(0.02)\text{Ratio}$$

2. CWC:

$$\text{Math}=60.19 + 1.20^{**}(\text{Homework}-X_{.j}) + 3.84^{**}(\text{SES}-X_{.j}) - -7.12^{**}(0.40)\text{Public} - 0.29^{**}\ (0.02)\text{Ratio}$$

3. CWC:

$$\text{Math}=47.53 + 1.20^{**}(\text{Homework}-X_{.j}) + 3.84^{**}(\text{SES}-X_{.j}) + +0.60^{*}(0.29)\text{Public} - 0.06^{**}(0.02)\text{Ratio} + 1.61\ (0.20)\ \text{Homework-Mean} + 7.98\ (0.25)\ \text{SES-Mean}$$

Note: the standard errors are no longer reported for the micro level predictors. The reason is that nor the value of the coefficients, nor their standard errors changed noticeably from RS to CWC, as was expected (see Kreft et al. 1995).
** is significant at p=0.01 level

The analyses so far have shown that users of multilevel models should carefully choose the way they handle the data. They should be made aware of the different effects of centering and the effects of omitting essential information, such as means, from the data. In general it can be expected that macro predictors are affected by deleting means from the model, especially when these means are so clearly related to the macro level variables, as is the case with mean SES and the Private Sector. The question what the correct model is, and what the correct sector effect is, depends on the purpose of the analysis, or the scenario. The answer to the question who is better, the private sector or the public sector, cannot be answered, because multilevel models are not causal models, although latter is suggested by the running head of a paper presented by Lee and Bryk (1988). What conclusions to report out of these analyses will depend on who asks the questions, and what the consequences are of such an answer.

Mean SES and the Private sector, A Summary

Knowing that the Means of the predictors SES and Homework are related to Private/Public sector, and knowing that high SES parents are over represented in the Private school it is not surprising that leaving those two means out of the CWC analysis will enhance the positive effect of the Private sector. Correlations for the NELS88 data are, -0.40 between Public and SES. Homework shows also a negative relation with the Public sector ($r=-0.14$), while Homework and SES correlate $r=0.21$ with each other. Individual SES is highly correlated with mean SES ($r=0.65$).

In the next table, Table 5 the only model so far that shows a positive effect in favor of the public sector is model 5. The other models show various (negative) effects for the public sector. This effect is strongest in model 2 and 4, where the variable SES and Homework are used in deviation of the mean, while both means are omitted from the model. In model 2 the effect of Public is -5.42, in Model 4 the effect is -3.24. The same models, but with Means reintroduced, show either a positive effect for Public (model 3) or a non-significant effect for Public (model 5). Comparing model 1 with model 6, model 2 with model 4, and model 3 with model 5 it is clear that adding Ratio to the model lowers the effect of Sector. It may mean that teacher/student ratio in the private school may be one of the secrets of their success, if that success is present. Comparing the results of Model 4 and 5 shows that not reintroducing the means has the similar effect on the coefficient for Ratio, as it has on the coefficient for Public. This effect is three times larger (from -0.6- to 0.18) in the model without means, than in the model with means. The difference in deviance between the two models is 153832 - 152998 = 834, showing that model 5 has a much better fit. In this case the RS model (model 6) has the best fit. Again we see that the coefficients for Ratio and Public enhanced when using a RS model.

Table 5 Summary of the CWC effect on macro level predictors

1. Partly CWC, with Public:
Math=52.24 + 1.25 Homework + 3.84 (SES–$X_{.j}$) - 4.97(0.37)Public

2. CWC, without Means, with Public:
Math=55.06 + 1.18 (Homework–$X_{.j}$) + 3.84 (SES–$X_{.j}$) - 5.42(0.36)Public

3. CWC, with Means, and Public:
Math=47.53 + 1.20 (Homework–$X_{.j}$) + 3.85 (SES–$X_{.j}$) -
-1.65 (0.20) Homework-Mean + 8.14 (0.25) SES-Mean + 0.62 (0.28)Public

4. CWC, without Means, with Public and Ratio: Deviance 152998
Math=49.49 + 1.17(Homework–$X_{.j}$) + 3.83 (SES–$X_{.j}$) -
- 0.18 (0.02) Ratio - 3.24(0.39)Public

5. CWC, with Means, Public and Ratio: Deviance 153832
Math=49.35 + 1.20 (Homework–$X_{.j}$) + 3.84 (SES–$X_{.j}$) -
-1.61 (0.20) Homework-Mean + 7.90 (0.25) SES-Mean -
-0.6 (0.02) Ratio - 0.39 (0.29)Public

6. RS with Public and Ratio: Deviance 153292
Math=53.13 + 1.24 Homework + 4.34 SES -
-0.16 (0.02) Ratio - 2.38 (0.37) Public

Using centered predictors versus raw scores is clearly a different scenario. It shows that some manipulation with the variables can reverse a favorable effect for the private sector into non favorable effect. All without adding any new variables but only manipulating variables by putting them in deviation from their mean, while not reintroducing that mean back in the model. Table 5 shows that in a summary.

Same model, with interaction effects

In the next analyses cross level interactions are added to the model, between two macro predictors Public and Ratio and two micro predictors, Homework and SES. Last variable, SES, is not allowed to interact with the macro-level variables in the RS model, because the slope for SES is not significantly random in that model, as shown earlier. I think the interaction between SES deviance score and Public, as in Raudenbush and Bryk (1986), should for that reason not be fitted too. For illustrations sake I do fit this interaction, but only in the CWC models. The variable Homework behaved differently. The slope for Homework was significantly random in both cases, Homework as raw score and Homework in deviation score. The effects of cross level interactions are summarized in Table 6.

Table 6. Cross level interactions in RS and CWC

1.RS: Deviance: 153272

$$\text{Math}=54 + 0.86(0.10)\text{Homework} + 3.84\ (\text{SES}-X_{.j})\ -3.35\ (0.37)\text{Public}$$
$$-0.16\ (0.02)\ \text{Ratio} + 0.48\ (0.11)\ \text{Public x Homework}$$

2. CWC with Means: Deviance: 152964

$$\text{Math}=49.57 + 0.82(0.10)(\text{Homework}-X_{.j}) + 2.97(0.26)\ (\text{SES}-X_{.j})\ - 0.60(0.38)\text{Public}+$$
$$- 0.06\ (0.02)\text{Ratio} + 1.30\ (0.21)\ \text{Homework-Mean} + 7.97\ (0.26)\ \text{SES-Mean} -$$
$$+ 0.48\ (0.11)\ \text{Public x Homework} + 1.03\ (0.28)\ \text{Public x SES}$$

3. CWC without Means: Deviance: 153808

$$\text{Math}=60.19 + 0.48(0.10)(\text{Homework}-X_{.j}) + 2.94(0.26)\ (\text{SES}-X_{.j})\ -7.12\ (0.40)\ \text{Public}$$
$$- 0.29\ (0.03)\ \text{Ratio} + 0.88\ (0.11)\ \text{Public x Homework} + 1.03\ (0.28)\ \text{Public x SES}$$

In Table 6, it shows that in all models the interactions are significant, where the public sector enhances the effect of homework and SES on math achievement in comparison to the private sector. Similar findings are in Raudenbush and Bryk, who labeled this "the egalitarian effect of the private sector" (l.c. p.11). Again, I want to stress that talking about the slope of SES, that this slope behaved differently in RS versus the CWC model. Besides it is not correct to talk in causal terms and state that the private school is more egalitarian. Nothing in RC models allows researchers to state causal relationships. This is even more clear if we remember the dramatic changes that take places if the data is manipulated by centering. Comparing the model without mean, models 3, with RS model and CWC with mean model (models 1 and 3) shows again that coefficients for Public and Ratio are strongest in model 3. Causal inferences based on one or the other model will lead two opposite causal inferences regarding the sector effect. The same is true for adding an interaction. Interactions are related to the main effects, and are bound to bounce around, as is clear from comparing the effects of Homework in deviation, and the cross level interaction between that Homework score and Public over the models 2 and 3. In model 2 Homework has a coefficient of 0.82 and the interaction scores 0.48. In the next model this balance is reversed, and the cross level interaction gets the benefit, with a value of 0.88 while Homework drops to 0.48, almost a complete reversal. A common occurrence when interactions are included in an analysis. Causal inferences about 'more egalitarian' Private sector, where Homework does not count as much as in the Public Sector, are again dangerous to make, as are the statements of that character in Raudenbush and Bryk. Correlations between interaction variables and one or both of the main variables introduces multi collinearity in the model, and hence bouncing beta's. In this data the correlation between the interaction variable Homework x Public and its main effects is r= 0.69 with Public, and r=0.71 with Homework Deviation Scores. The correlation of the interaction variable SES in Deviation and Public = 0.0 (n.s.), and 0.88 with SES in deviation of the mean (checking the same correlations using raw scores yield a similar pattern).

Is a Centered Model A better Between-Regression Model?

Centering predictors can cause flip flop effects of the macro level variables, as illustrated in the examples. Reevaluation is needed of the application of centering in RC modeling to predictors, since straightforward generalization from traditional regression to RC models is not correct (see Kreft et al. 1995). Centering of variables around the group mean has a history in education since Cronbach (1976) advocated such a model for the separation of student effects from school or class room effects, and it is to this discussion that Raudenbush and Bryk (1986, p.12) refer when they express their preference for group mean centering; "..if we center the numerical variables around their respective school means, the intercept captures the mean achievement level in each school. Thus the between-group equation for the intercept represents the regression of mean achievement or school level factors as equivalent to using the school as the unit of analysis. Further as noted previously, the within group model represents the pooled within school regression of achievement on student level characteristics. In this model the analyst can decompose effects into their within- and between-school components, as advocated by Cronbach." I think it is not quit that simple, and centering or not centering should be applied according to the goals and purposes of the research. The analyses given in this paper can serve as an example.

If the scenario is to find school effects for policy actions and/or to consult government agencies I could conclude that better math achievements would be obtained when the teacher/student ratio was lowered, or that vouchers for the private sector should be issued to all parents. My advice would be that publications that advocate that private schools do a better job than public schools should first be reexamined before taken serious action.

Based on the NELS88 data set analyzed in different ways I come to the conclusion that homework makes a difference, as does the level of the students' home environment, as defined by parents income and educational level. The public school is less effective than the private school, but largely as the result of mean SES and mean Homework levels of the school. I based the conclusion on the fact that the sector effect disappears when the raw scores are replaced by deviation scores and their respective means. The teacher/student ratio seem to be fairly stable and less influenced by the fact that either raw scores or deviance scores are used. All analyses show the same effect, which is that the lower the teacher student ratio the better the results in math, even after correction for SES, Homework, SES mean, Homework mean, Public sector and the interactions between Public and, respectively, Homework and SES.

## DISCUSSION

For a better understanding of what centering a description is given of the relationship between the estimates of the coefficients in a linear model that contains the individual level elements $X_{ij}$ versus $(X_{ij} - X_{.j})$ together with the aggregated context level element $X_{.j}$. First the coefficient for the prediction of $Y_{ij}$ from $X_{ij}$ across all data from all contexts is defined as $b_T$. Where T means Total. Making use of the fact that $b_T$ is a composite of the between-group regression $b_B$ estimate and the pooled within group regression estimate $b_W$ in the following way:

$$b_T = \eta^2 b_B + (1-\eta^2) b_W \qquad \text{where}$$

$\eta^2$ is the explained proportion of the variance in the dependent variable between classes. Relations between coefficients in RS and CWC can be defined more properly now. For instance in the CWC model, the effect $(X_{ij} - X_{.j})$, corrected for the effect $X_{.j}$ is equal to $b_W$, while the effect of $X_{.j}$ corrected for $(X_{ij} - X_{.j})$ is equal to $b_B$. The same effect for $X_{.j}$, but corrected for the raw score effect of $X_{ij}$ (as in the RS model) is equal to $(b_B - b_W)$ (see e.g. Duncan et al. 1966). This shows that centering around the context mean changes the definition of the context effect c. .he group mean from $(b_B - b_W)$ to $b_B$. In fixed effects models group mean centering separates and orthogonalizes the between variation from the within groups variation.

In this paper I have shown how group mean centering in RC models changes parameter estimates in a way that does not directly relate to the RS model, unless the group mean is reintroduced. Centering within context is, contrary to centering around the grand mean, not without consequences, and theories are needed to support this approach to model specification and analysis. In the literature I found two different reasons for centering of predictors, one mathematical, and one theoretical.

Mathematically Driven Choice of Model

Is there a statistically 'correct' choice among RS, and CWC? The answer is 'no', because the models are all correct, and from the point of view of model estimation, this question is only one of computational ease. If predictor variables have widely differing scales, e.g. SAT scores with a mean of 500 and a standard deviation of 100 versus a grade point average scale with a range of four, centering is called for. Since scales in psychology are in the main arbitrary, rescaling predictors to approximately equal locations and variances prior to analysis is typically possible and desirable. Thus, except for the computational argument that leads equally to the choice of grand mean centering and CWC, the choice between models cannot be based on statistical arguments.

Theory Driven Choice of Model

Group mean centering in random coefficient models is a topic discussed in a technical way in a recent publication (Kreft, DeLeeuw and Aikin, 1995), who reach the conclusion that centering predictors is fitting another model and is not comparable with raw score models. The discussion of centering within context is in general not new (Cronbach, 1979, Burstein, 1980), but the question is, if its use in RC models can be advocated indiscriminately. I have shown with examples that results differ. Researchers need conceptual reasons for using either raw scores or centered scores. Reasons or rules for centering cannot be given in general. Each research has its own scenario, and based on that scenario the researcher has to make a choice. All that is done in this paper is making people aware of consequences of some of the choices.

For instance, using centered scores (without context means reintroduced at the macro-level) yields a worse fitting model because group variation is eliminated from the micro-predictors, and totally eliminated from the model in cases where the group mean is not reintroduced. A researcher choosing last type of centering is either implicitly or explicitly assuming that such between class variation is not meaningful.

In analysis of variance, where scores are also put in deviation of the group mean, it is assumed that the effect measured is conditional on the group to which that individual belongs. As in my analysis, for instance, being a student in the public or the private sector. People are treated as reacting equivalently, depending on their deviation score, even when means are quite different for different groups. This notion is based on the assumption of experimental models that apply random assignment to groups; group mean deviations at pretest are treated as random fluctuations from that mean, while at post test the group means reflect the treatment effect. In quasi-experimental designs, where group membership is structural and meaningful, using centering within groups instead of raw scores removes meaningful differences between observations. Centering is based on the assumption that a high scoring student, but below the group mean, in high achieving schools is equivalent to a very low scoring student equally distant from the group mean in low achieving schools. This is the same as assuming that students behave more in relation to their peers, than to their own aptitude for math.

It seems that the choice between a CWC model with context means added as macro-predictors, and a raw score model, should be made carefully, and based on theoretical rather than technical considerations. For instance, to use the group mean as a predictor a theoretically sound explanation for the relationship of the context means to the outcome is needed. The relation between math achievement and homework across sectors provides a good example. The average homework of a school reflects to some extent parental and school climate values, as does the average SES. The relationship of Homework to math achievement reflects the importance of the effort of each student within schools. Or is it that the amount of homework should not be measured in real hours, but in hours in comparison to the average number of hours of the school, or school class? This question refers to the need for one model or for two separate models? Two models that account for the homework/math achievement relationship within versus between schools. The answer is likely two models, in instances where we expect that different forces operate on the amount of homework made by students. One force on the hours dedicated to homework within the school, like peer comparison, as measured by the deviation score. One force between schools, like school climate, measured by the proxy "means hours of homework".

For the use of centering I see two possible scenarios. First, if preexisting mean differences between contexts can be defined centering could be chosen. The choice is still between a centering model with or without adding the context means back to the model at the macro-level. Throwing away the context mean is throwing away important information. Second, if there exist a distinct theory for the relationship of the centered micro-predictor to the outcome and another distinct theory for the relationship of context means to the outcome. The centered model with context means reintroduced at the macro-level can be employed as the final test of such a theory. But this model should not be chosen when there is no reason to expect the macro- versus micro- distinction with regard to a single predictor to be meaningful.

I see the past tense in Raudenbush and Bryk's concluding comments in their 1986 article as a little optimistic. They wrote: "Research on school effects has been plagued by both methodological and conceptual problems (l.c. 1986, p.15). It still is. But RC modeling does offer many more ways to analyze the same dataset, forcing researchers to conceptually rethink their models, and enhancing new developments. Therefore I agree with the part that follows the citation in Raudenbush and Bryk's 1986 paper, where they mention that RC models are a promising development, that it greatly expands the range of methods for investigating schools, thereby expanding conceptualization.

For instance, statements made in the early days before RC modeling by Coleman, Hoffer and Kilgore (1982, p.196) that "there is a tendency to converge over time among students from different backgrounds in Catholic schools, and a tendency toward divergence in the public school", and similar findings by Raudenbush and Bryk (1986) have to be reexamined. My analyses do not support a voucher system, as proposed by Ronald Reagan (1983), president of the United States, to solve the "Nation at Risk" problems. I think school effectiveness research is still too much plagued by methodological problems to support any conclusions, such as that private schools do better than public schools. Looking at the result of my raw score analysis, where the slope of SES predicting Math is not different over schools, and finding that the main effect of sector disappears when controlling for mean SES and mean Homework in the CWC model, I conclude that centering has consequences that need attention and explanation. Some of these consequences may be artifacts.

## REFERENCES.

Anderson, C.S. (1982). The search for school climate; a review of the research. *Review of Educational Research*, 52, 3, 368-420.

Bryk , A.S. and S.W. Raudenbush (1992). *Hierarchical Linear Models. Applications and Data Analysis Methods.* Newbury Park: Sage

Bryk, Anthony S, Raudenbush, Stephen W, Seltzer Michael and Congdon Richard T (1988) *An Introduction to HLM: Computer Program and Users' Guide.*

Burstein, L., Linn, R.L., and Capell, F.J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, 3, 347-383.

Boyd, L.H. and Iversen, G.R. (1979). *Contextual analysis:concepts and statistical techniques.* Belmont, CA: Wadsworth.

Coleman, J.S. Hoffer, Th, and Kilgore, S. (1982). Cognitive Outcomes in Public and Private Schools. Sociology of Education, 55, 162-182.

Duncan, O.D., Curzort, R.P. and Duncan, R.P. (1966). *Statistical Geography: Problems in Analyzing Areal Data.* Glencoe Free Press.

Kreft, Ita G.G. and Jan De Leeuw and Leona S Aiken (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research* ,

Kreft, Ita G.G., Jan de Leeuw and Rien van der Leeden (1994) Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML2 and VARCL. *The American Statistician*, 48, 4, p 324-335.

Kreft, Ita G.G., De Leeuw, Jan and Kyung-Sung Kim (1990). *Comparing four different statistical packages for hierarchical linear regression GENMOD, HLM, ML2 and VARCL, CSE Technical Report #311*, Center for the Study of Evaluation. UCLA (see ERIC tracking # TMO17778).

Lee, Valerie E. and Anthony S. Bryk (1988). *A Multilevel Causal Model of the Social Distribution of High School Acheivement.* Presented at the AERA

Lee, Valerie E. and Anthony S. Bryk (1991) In: Raudenbush, Stephen W. and J. Douglas Willms (Eds.). *Review of Schools, Classrooms and Pupils, International Studies of Schooling from a Multilevel Perspective.*, SanDiego: Academic

Longford, Nicholas T. (1990) VARCL. *Software for Variance Component Analysis of Data with Nested Random Effects* (Maximum Likelihood), Educational Testing Service, Princeton, NJ.

Rasbash, J,., Prosser, R., and Goldstein, H. (1989). ML3. *Software for Three-Level Analysis. Supplement.* London: Institute of Education, University of London.

Raudenbush, Stephen W. and Anthony S. Bryk (1986). A Hierarchical Model for Studying School Effects. *Sociology of Education*, 59, 1-17

Raudenbush, Stephen W. and J. Douglas Willms (Eds.) (1991). *Review of Schools, Classrooms and Pupils, International Studies of Schooling from a Multilevel Perspective*, SanDiego: Academic

Reagan, Ronald (1983) A Nation at Risk. Presidential address