KM 24(1987)
pag 137-148

The SeeSaw Effect: A Multilevel Problem?
A reanalysis of some findings of Hox and De Leeuw

Ita Kreft & Edith de Leeuw *

ABSTRACT

Studies of schooleffectiveness often use measures of association, such as regression weights and correlation coefficients. These statistics are used to estimate the size of the change or 'effect' that would occur in one variable (for example reading ability) given a measured change in another variable (for example sex and sex ratio). In this paper we explore the limitations of regression coefficients for use in a contextual analysis, in which both individual and contextual variables are included as independent variables. In our example 'individual sex' and a context variable: 'sex ratio of the schoolclass' are regressors, and reading ability is the dependent variable. Our conclusion is that researchers should be careful making interpretations of effects from multiple regression analysis, when dealing with aggregate data. Even in the case (as in our example) when individual and contextual variables are made orthogonal to avoid multicollinearity, interpretation of the effects of the aggregate variable is problematical.

Keywords: EDUCATIONAL RESEARCH; MULTI LEVEL RESEARCH; CONTEXTUAL ANALYSIS; HIERARCHICAL STRUCTURED DATA; ECOLOGICAL FALLACY; ROBINSON EFFECT.

The authors would like to thank prof. dr. J. de Leeuw and dr. J.Hox for taking part in the discussion, and dr. J.C. van der Wolf for his permission to use his exciting data.

----

* Department of Education, University of Amsterdam
Prinsengracht 227, 1015 DT Amsterdam, Holland
Telephone: 020-5252641

1. INTRODUCTION

The first discussion of the problems encountered in research which deals with units of more than one level are found in the sociological literature (c.f. Lazarsfeld, 1959; Lazarsfeld & Menzel, 1961). The analysis of multilevel data has also received considerable attention in the field of educational research (Oosthoek & Van der Eeden, 1984). Especially school effectiveness studies, where the importance of school (class) level on the dependent variable (pupil achievement) is investigated, have been the object of much debate. The best known example is the study of Coleman, Cambell, Hobson, McFarland, Mood, Weifeld, & York (1966), and the discussions around Coleman's results (Jencks, 1973; Mosteller & Moynihan, 1972; Mayeske, Okada, Beaton, 1973).

Recently new models have been proposed to analyze hierarchically structured data such as data on pupils within classes within schools (Aitkin & Longford, 1986; De Leeuw & Kreft, 1986; Mason, Wong, & Entwisle, 1983). These models, known as variance component or random coefficient models, are still in the developing stage. Therefore, researchers have no other option but to use the linear regression model with fixed coefficients.

One of the difficulties one can encounter in interpreting the effects of variables from different levels in one single equation is the 'Robinson effect', which can lead to the 'ecological fallacy'. This effect, first described by Robinson (1950), occurs when relations, determined between aggregated data, are indiscriminantly translated to the individual level (Van der Eeden & Huttner, 1982). To illustrate the dangers of these 'indiscriminate translations' two examples of ecological fallacies are given.

The first example is Robinson's original one. He showed that the correlation between percentage blacks and percentage illiterates was 0.80 at the state level, while the correlation at the individual level between being black and illiterate was 0.19. The second example shows even more clearly the error one makes in committing the ecological fallacy. This example is given by Hox & De Leeuw (1986). They point out that during the last elections an extremely right wing, racist party (Centrum partij) received relatively more votes in districts in which many foreign laborers live. The conclusion that foreign laborers voted for the Centrum partij can proved to be wrong. At that time in the Netherlands foreign laborers were not allowed to vote! For a further discussion of the ecological fallacy see also Kreft & Van den Eeden (1985).

In their (contextual) analysis of Dutch educational data Hox & De Leeuw (1986) found some extreme cases of the 'Robinson effect'. In a multiple regression analysis the beta-weights at individual (pupil) level did not only differ from those for the same variable at aggregate (class) level, they even changed sign! This was the case for sex of pupil as a predictor of reading ability. Hox and De Leeuw christened this unexpected effect the 'See-saw effect'. An interpretation could be that although girls have better reading results than boys, in classes with a high percentage of girls the teachers overall rating of reading ability is lower (frog-pond effect, Burstein, 1980). This plausible interpretation about a change of meaning of the variable (sex) with a change of level is however not the appropriate one. At least not in this case, as we will show in this paper.

2. THE DATA

2.1. Data collection

In the years 1983 and 1984 data have been collected in six primary schools in Amsterdam. The data collection was part of the research project "Preventie van school- en leerproblematiek (prevention of school and learning problems)" of the Department of Education (Orthopedagogisch Instituut) of the University of Amsterdam. A total of 681 pupils from 29 classes, and their teachers, completed the tests and questionnaires. In October 1983 the pupils completed a sociometric test, a test measuring their attitude towards school, a questionnaire about deviant behaviour in the school, and a standardized (CITO) test for reading. The teachers were, among other things, asked about their coping behavior in problem situations and their job satisfaction. In June 1984, at the end of the 'school year', the teachers were asked to rate their pupils on reading and arithmetic. In addition the teacher was asked to indicate whether a pupil would be promoted to the next class, and whether a pupil would do better in a special school for children with learning disabilities or behavioral problems (BUO). For a complete description of the questionnaires used, see Van der Wolf (1984). Some psychometric analyses were reported earlier in Van der Wolf, Hox & De Leeuw (1985).

2.2. Method used by Hox & De Leeuw

Using regression analysis Hox & De Leeuw (1986) tried to predict the achievement of pupils in various fields. The dependent variables were 1. rating of reading ability, 2. rating of arithmetic ability, 3. promotion to the next class, 4. desirability of transfer to a special school. These four dependent variables were all measured at individual (pupil) level in June 1984.

The independent variables were measured in October 1983, both at individual (pupil) level, and at class level (teacher characteristics). The following pupil characteristics were used as independent variables: sex of pupil, number of times a pupil was not promoted to the next class, membership of minority group, physical disability, and test score for reading ability. For each pupil characteristic three new variables were defined: the characteristic at school level, the characteristic at class level, and the characteristic at individual level. These characteristics were computed by transforming each raw pupil score into three uncorrelated components: school mean, deviation of class mean from school mean, and deviation of raw pupil score from class mean (cf. Cronbach & Webb, 1975; Harnqvist, 1978). In other words, if X_{ijk} is the raw score of pupil i in class j on school **k**, then the predictor at the individual level is $X_{ijk} - X_{.jk}$. The predictor at class level is then $X_{.jk} - X_{..k}$, and the predictor at school level is $X_{..k}$. The following teacher characteristics were used as independent variables: sex of teacher, attendance of special courses in education, years of teaching experience, number of schools they teached at, and two composite scores based on factor analysis of teacher variables such as coping, job satisfaction etc. (see Van der Wolf, 1984). For each teacher characteristic two new variables were defined: one at school level, and one at class level. This was done by transforming each raw teacher score into two orthogonal components: school means and deviation of raw teacher score from school means.

The order of entering variables in the regression equation was as follows: 1. all pupil variables at individual level, 2. all pupil variables at class level, 3. all teacher variables at class level, 4. all pupil variables at school level, 5. all teacher variables at school level. In this reanalysis we used the same variables as Hox & De Leeuw.

3. THE CASE IN QUESTION

3.1. Some findings of Hox & De Leeuw

In some of the regression equations, as reported by Hox & DeLeeuw, the predictor sex of pupil, among others, behaved rather strangely. At the individual level the beta-weight of sex was positive, at the class level the beta-weight was negative. See Table 1 for the coefficients given by Hox & De Leeuw. The corresponding multiple correlation was R = 0.59.

Table 1. Prediction of Rating of Reading Ability.

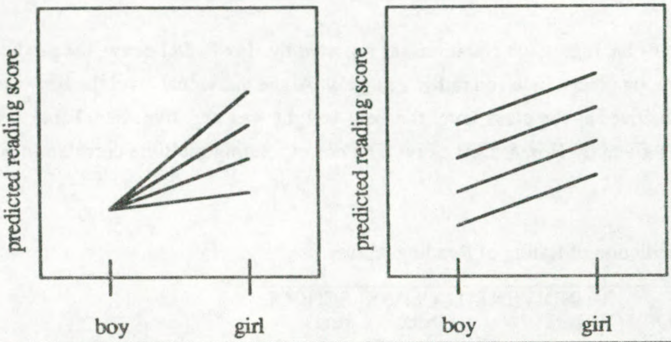
PREDICTOR	INDIVIDUAL beta	CLASS beta	SCHOOL beta
sex pupil	.10*	-.24*	-
not promoted	-.08	-.05	-
minority	-.09*	.01	-
disability	-.09*	-.08	-
reading test	.43*	.06	-
sex teacher	n.a.	-.04	-
special courses	n.a.	.02	-
experience	n.a.	-.07	-
previous schools	n.a.	.06	-
composite 1	n.a.	-.06	-.14*
composite 2	n.a.	-.07	-

* means significant ($\alpha = .05$), n.a. means not available

This 'seesaw-effect' of sex of pupil was totally unexpected. Of course, it is possible to find post-hoc explanations. For instance, the finding that girls score higher on reading ability than boys, but classes with a majority of girls do worse, could be consistent with:

1. In classes where boys are in the majority girls do better, or are rated higher by their teacher. In terms of regression analysis: classes have the same intercept, but different slopes (see Figure 1).

2. In classes where girls are in the majority both boys and girls do worse, or are rated lower. Classes have the same slope but different intercepts (see Figure 2; and Boyd & Iversen, 1979, p.129).



Figures 1 and 2: regression lines for various school classes.

But using post-hoc explanations is rather dangerous (cf. Cliff, 1983). On examining the data we actually find only sampling variability between slopes and/or intercepts in the different classes. As we will show in the next section both post-hoc explanations stated above are inconsistent with the data.

3.2. Redefinition of the problem

For the clarity of the argument we will redefine the problem and utilise a simplified version of the equation of Hox & De Leeuw, using only the individual level and the class level, and sex of pupil as the only independent variable. This results in two regressors: an individual predictor 'sex' X_{ij} (i=pupil, j=class), taking only the two values 1=boy and 2=girl, and a class level predictor 'sex ratio' $X_{.j}$. The dependent variable is again reading ability as rated by the teacher. Complete data were available for N = 513 pupils.

The Cronbach procedure leads to the following model, with two (unstandardized) regression weights:

$$Y_{ij} - Y_{..} = b_1(X_{ij} - X_{.j}) + b_2(X_{.j} - X_{..}) + e_{ij}.$$

In the next section, and the appendices, we will show that the 'seesaw-effect' described by Hox and de Leeuw is not a conceptual problem but an artifact of the linear model used, and of the corresponding test of significance. Therefore, the conclusion that there is a different effect on reading score between the predictor sex at individual and class level is questionable.

4. ANALYSIS AND RESULTS

The total variance of the pupil characteristic sex and reading scores can be divided into the variance between classes, and the variance between pupils within classes. Here we have to remember that the between school variance is computed by taking the number of pupils of the schools into account. Thus the between variance is not the same thing as the variance of the school means, but a weighted version of this schoolmean variation. This partitioning of the variances and covariances, given in Table 2, can be used to compute the required regression coefficients.

When we inspect these matrices we see that most of the variance is within classes. Remember that the correlation ratio η^2 is the proportion of variance that is between classes. For sex $\eta^2 = .03$, and thus the individual level accounts for 97% of the total variance. The class level only accounts for 3% . In other words a very small part of the total variation is at class level. This means that classes are very much alike for this variable.

Table 2. Partitioning of 'reading ability' and 'sex of pupil'.

	Total		Within		Between	
	sex	reading	sex	reading	sex	reading
sex total	.250	.022				
reading total	.022	.535				
sex within			.243	.030		
reading within			.030	.460		
sex between					.006	-.008
reading between					-.008	.075

If we calculate the percentage of variance between classes for reading ability we find a value of $\eta^2 = .075/.535 = .15$, much larger than the correlation ratio for sex. This means that 85% of the variation in reading scores is within classes. The total covariance between sex and reading is small (.022), the covariance within classes larger (.034), which results in a negative covariance at between class level (-.008).

As a consequence of the fact that the covariance within schools is larger than the total covariance, the regression coefficients within and between classes have opposite signs (see appendix B for a proof). Calculating the unstandardized regression weights from Table 2 gives the result that $b_1 = .030/.243 = .123$ for the individual variable, and $b_2 = -.008/.006 = -1.46$ for the class level variable. The significance test, based on the standard multiple regression model, declares both values to be significant at a level of 0.05. But we have reason to doubt the significance of the regression weight at class level, especially since it is based on very small percentages of explained variance. Because there is more then one source of error (at individual, and at class level) the standard significance test is in this case not a valid one (cf. De Leeuw & Kreft, 1986).

5. CONCLUSION

It is clear by now that the regression weight at the class level in the fixed effect model has no reliable interpretation as a context effect. The results are unpredictable and can be meaningless, because they depend entirely on using a test of significance, which is based on erroneous assumptions about the error terms. The interpretation of standard errors and hypothesis tests leans more heavily on distributional assumptions than point estimators do. For instance, the assumption of independence of the observations. Pupils share a common experience when taught by the same teacher in the same class. As a consequence of the hierarchical structure of 'school research data' the assumption of independent observations is violated. Without taking the more complex error structure in these data into consideration, it becomes impossible to test the regression coefficients for significance, and interprete those in the usual way. School effectiveness studies must therefore move away from the means to means approach and from mixing different levels in one model to more appropriate models for hierarchically structured data.

Although the authors are not aware of any analytic or Monte Carlo studies on this subject, the results of Hox & De Leeuw strongly suggest that the standard significance test of regression weights, as applied to multilevel problems, is not robust. In particular, the operative alpha level is much higher then the nominal alpha level.

Regrettably there are only few statistical programs, that address the multilevel problem. Ideally, what is needed is a well described, readily available computer program or statistical package with proper test of significance. We are experimentating with the technique proposed by Aitkin and Longford (1986), which is implemented in the VARCL program. In the mean time researchers have to be very careful when analyzing multilevel data. As a general strategy we propose:

1. To be very wary of post-hoc explanations of multi-level data.
2. To inspect (co)variance matrices on different levels and to compare these with the overall (co)variance matrix. To produce a printout of a variance-covariance matrix is simple; for instance 'statistic 2' of the SPSS-procedure 'Reliability' will do the trick.
3. Where ever possible to estimate the standard errors with a Monte Carlo procedure, for instance 'bootstrapping' (Efron, 1979) or 'jack-knifing' (Mosteller & Tukey, 1968).

Appendix A: An identity for covariance matrices and regression weights.

Suppose we have two variables x and y, in a design which has two levels. Then both the data matrix and the covariance matrix can be divided into a between-groups matrix and a within-groups matrix. For the covariance matrix the following equation holds

$$C(x_T,y_T) = C(x_B,y_B) + C(x_W,y_W). \tag{A1}$$

In the technique of Cronbach and Webb (1975), as used by Hox and De Leeuw (1986), the covariance matrices $C(y_T,x_B)$ and $C(y_T,x_W)$ are also used. From

$$C(y_T,x_B) = C(y_{B+W},x_B) = C(y_B,x_B) + C(y_W,x_B), \tag{A2}$$

and

$$C(y_W,x_B) = C(y_B,x_W) = C(x_B,x_W) = 0 \tag{A3}$$

follows

$$C(y_T,x_B) = C(y_B,x_B). \tag{A4}$$

In the same way

$$C(y_T,x_W) = C(y_W,x_W). \tag{A5}$$

As a consequence the regression weights (with x_B and x_W as predictors) can be written as

$$b(y_T,x_B) = C(y_B,x_B)/V(x_B), \tag{A6}$$

$$b(y_T,x_W) = C(y_W,x_W)/V(x_W). \tag{A7}$$

Appendix B: Sufficient conditions for the 'seesaw-effect'.

Let us write

$$b(y_B,x_B) = C(y_B,x_B)/V(x_B) = \{C(y_T,x_T) - C(y_W,x_W)\}/V(x_B) =$$
$$= \{b(y_T,x_T)V(x_T) - b(y_W,x_W)V(x_W)\}/V(x_B). \tag{B1}$$

Dividing numerator and denominator by $V(x_T)$, and defining $\eta^2(x) = V(x_B)/V(x_T)$, gives

$$b(y_B,x_B) = \{b(y_T,x_T) - b(y_W,x_W)(1 - \eta^2(x))\}/\eta^2(x). \tag{B2}$$

If $b(y_T,x_T) \approx 0$ it follows from (B2) that

$$b(y_B,x_B) = - b(y_W,x_W)(1 - \eta^2(x))/\eta^2(x). \tag{B3}$$

Thus $b(y_B,x_B)$ and $b(y_W,x_W)$ will have opposite signs, and $b(y_B,x_B)$ will be much larger if $\eta^2(x)$ is small.

If $\eta^2(x)$ is very small, we can find other sufficient conditions for the seesaw-effect. We have the approximation, from (B2),

$$b(y_B,x_B) \approx \{b(y_T,x_T) - b(y_W,x_W)\}/\eta^2(x). \tag{B4}$$

Thus if $b(y_T,x_T) > b(y_W,x_W)$ and $b(y_W,x_W) < 0$ we find opposite signs, and the same thing is true if $b(y_T,x_T) < b(y_W,x_W)$ and $b(y_W,x_W) > 0$. In our example, and presumably in many others with small $\eta^2(x)$, we find that total and within regression coefficients are approximately equal and positive, with total slightly smaller. This will make the between regression coefficients both large and negative.

6. REFERENCES

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. Journal of the Royal Statistical Society, 149, (part 1), 1-43.

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

Burstein, L. (1980). The analysis of multilevel data in educational research in evaluation. In: D. Berliner (Ed.), Review of Research in Education, Volume 8, (pp.158-233). Washington D.C.: American Educational Research Association.

Cliff, N. (1983). Some cautions concerning the application of causal modelling methods.

Multivariate Behavioral Research, 18,115-128.

Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPharland, J.,Mood, A.M., Weifeld, F.D. & York, R.L. (1966). Equality of Educational Opportunity. Washington D.C.: U.S. Department of Education and Welfare.

Cronbach, L.J., & Webb, N. (1975). Between class and within class effects in a reported aptitude x treatment interaction: Reanalysis of a study by G.L. Anderson. Journal of Educational psychology, 67, 717-724.

De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. Journal of Educational Statistics, 11, 1, 57-85.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1-26.

Harnqvist, K. (1978). Primary mental abilities at collective and individual levels. Journal of Educational Psychology, 70, 706-716.

Hox, J.J. & De Leeuw, E.D. (1986). De invloed van individuele en contextuele kenmerken op schoolbeleving en leerprestatie. In: J.C. Van der Wolf & J.J. Hox (Eds.), Kwaliteit van onderwijs in het geding. Lisse: Swets & Zeitlinger.

Jencks, C. (1973). Inequality, a reassessment of the effect of family and schooling in America. New York: Basic Books.

Kreft, G.G., & Van den Eeden, P. (1985). Valkuilen bij het proefschrift van L. van Marwijk Kooy, en hoe deze uit de weg te gaan. In : D.P. Kroes (Ed.), Studiebijeenkomst over de dissertatie " Scholen verschillen". Amsterdam: SISWO.

Lazarsfeld, P.F. (1959). Problems in methodology. In: R.K. Merton, L. Broom & L.S. Cottrell (Eds.), Sociology Today. New York: Basic Books.

Lazarsfeld P.F., & Menzel, H. (1961). On the relation between individual and collective properties. In: A. Etzioni (Ed.), Complex Organizations. New York: Holt, Rhinehart & Winston.

Mason, W.M., Wong, G.Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In : S. Leinhardt (Ed.), Sociological Methodology 1983. San Francisco: Jossey Bass.

Mayeske, G.W., Okada, T., & Beaton, A.E. jr. (1973). A study of attitude toward life of our nation's students. Washington D.C.: Government Printing Office.

Mosteller, F., & Moynihan, D.P. (1972). On equality of educational opportunity. New

York: Vintage Books.

Mosteller, F., & Tukey, J.W. (1968). Data analysis, including statistics. In: G. Lindzey & E. Aronson (Eds.), Handbook of social psychology, vol. 2. Reading: Addison-Wesley.

Oosthoek H., & Van der Eeden, P. (Eds.) (1984). Education from the multilevel perspective: models, methodology, and empirical findings. New York: Gordon & Breach.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, 15, 351-357.

Van den Eeden, P., & Huttner, H.J.M. (1982). Multilevel research, Beverly Hills: Sage.

Van der Wolf, J.C. (1984). Schooluitval, een empirisch onderzoek naar de samenhang tussen schoolinterne factoren en schooluitval in het regulier onderwijs. Lisse: Swets & Zeitlinger.

Van der Wolf, J.C., Hox, J.J. & De Leeuw, E.D. (1985). Schoolbeleving en schoolprestatie; een onderzoek naar mogelijke determinanten van schooluitval naar het buitengewoon onderwijs. In: K.B. Koster & N.J.M. Kwantes (Eds.), Schoolbegeleiding en leerproblemen. Lisse: Swets & Zeitlinger.

Ontvangen: 02-10-1986
Geaccepteerd: 13-01-1987