

## Models for Value-added Investigations of Teaching Styles Data

Neil H. Spencer  
*University of Hertfordshire*

*Abstract:* This paper considers models of educational data where a value-added analysis is required. These models are multilevel in nature and contain endogenous regressors. Multivariate models are considered so as to simultaneously model results from different subject areas. Path models and factor models are considered as types of model that can be used to overcome the problem of endogeneity. Estimation methods available in MLwiN and EQS are used. The use of a factor model with EQS is shown to give estimates of the effects of teaching styles that have smaller standard errors than any other method studied.

*Key words:* Educational data, multilevel modelling, structural equation modelling, value-added.

### 1. Introduction

In all statistical modelling, the choice of model is paramount. The use of a misspecified model may produce results that are misleading. In this context, multilevel modelling is now used routinely to analyse educational data as it models the hierarchical structures found in educational contexts. Section 2 gives further details, and also discusses the endogeneity issue for multilevel modelling raised by value-added investigations. It then goes on to look at ways of dealing with this issue that have been proposed in the past (Spencer and Fielding, 2002; Spencer, 2002) and the possibilities now available through the combining of multilevel modelling with structural equation modelling (SEM).

This paper extends the work done in the papers of Spencer and Fielding (2002) and Spencer (2002) by (i) considering a multivariate framework of models in which a number of current tests scores may be related to multiple prior test scores, rather than being restricted to univariate models; (ii) exploiting the new abilities of SEM packages to undertake multilevel modelling and of a multilevel modelling package to fit SEMs.

Section 3 of this paper introduces the Teaching Styles data of Bennett (1976). It also shows the results that were obtained by Bennett (1976) when investigating pupil progress for three different styles of teaching, those obtained by a reanalysis

by Aitkin, Anderson, and Hinde (1981), and another reanalysis by Spencer (2002). Section 4 then goes on to use a multivariate framework for the analysis of the data along with different ways of expressing the model as a path model and a factor model. Conclusions are given in section 5.

## **2. Modelling Considerations**

### **2.1 Multilevel modelling**

The vast majority of the research into educational processes takes place in the context of schooling. Pupils are the basic unit of investigation, and these are found grouped in classes which themselves can be considered to be grouped within the schools. These schools themselves can be considered to be grouped geographically and perhaps also administratively because of this. We thus have a hierarchical structure. Pupils can be considered the first level of the hierarchy with classes at the second level. Schools can be considered a third level with any geographical or administrative grouping being a fourth level. This hierarchical structure may vary depending on the particular circumstances of the investigation being undertaken. It is frequently the case that the school is considered as the highest level of the hierarchy, with the geographical/administrative level being disregarded, either because it is not considered to be relevant or because the study has only taken place within one geographical/administrative unit. In longitudinal studies where measurements are taken from pupils over time, the lowest level of the hierarchy can be considered to be the occasions at which the measurements are taken. These are effectively grouped within the individual pupils who now form a second level of the hierarchy. Class, school, etc. levels then build on top in the usual way.

The hierarchy may also be considered in a more complex way when teachers are considered. Teachers exist at the same level as the class, and a simple hierarchy with one teacher per class and one class per teacher would mean that the class and teacher are synonymous with each other. However, in practice a class may have more than one teacher and a teacher may teach more than one class. We thus have a cross-classified hierarchy where classes are grouped under more than one teacher and teachers are grouped under more than one class.

Traditional statistical techniques are often based on the assumption that the basic units of investigation are independent of each other. Because of the groupings that exist, this assumption is rarely tenable in educational research. Pupils that are in the same class cannot be considered independent of each other because they are subject to influences that are due to the class dynamics. Similarly, they will have been exposed to influences from the same teacher(s), and also to the influences from the same higher level units (e.g. schools). In longitudinal

studies where the lowest level of the hierarchy might be measurement occasions, it is clear that individual measurements taken from the same pupil cannot be considered independent of each other.

The lack of independence is an issue when it comes to the estimation of the parameters of a model. An appropriate way of proceeding is to recognise the groupings that exist in the data, and build these into the model. Thus, if it is accepted that a group of pupils may be experiencing similar influences as a result of being in the same class, then these influences should be included as part of the model. The same applies for any other potential influences such as those that are due to teacher, school or geographical/administrative unit. This could be done by including dummy variables for the grouping units as independent variables in the model. The associated parameter estimates would then represent the influences of the different classes, teachers, schools, etc. However, in many circumstances there are a large number of classes, teachers, schools, etc. involved in the study, and the above approach would necessitate the estimation of a large number of regression coefficients. The values of the vast majority of these coefficients would probably not be of particular interest in themselves, and thus a more efficient way of proceeding would be to regard the effects of the various units at a particular level (e.g. the different classes being the units at the class level) as coming from some common distribution. Thus, instead of having to estimate a separate regression coefficient for each unit at a level, only the parameters of the common distribution would need to be estimated.

The field of multilevel modelling has built on this approach to overcoming the lack of independence, and developed the flexibility to cope with cross-classified hierarchies, non-linear models and other extensions. Books by Goldstein (2003), Hox (2002), Raudenbush and Bryk (2002), Snijders and Bosker (1999) are amongst those that give good background information concerning multilevel modelling and go on to deal with various extensions.

## 2.2 Value-added investigations and endogeneity

In recent years attention in government education policy, particularly in the U.K., has been on the effects teachers and schools have on the progress made by pupils. It has been widely recognised that focusing on raw test results without adjusting for prior levels of attainment does not give an adequate picture of the educational processes at work and the quality of education delivered by the teachers and schools (e.g. Bird, Cox, Farewell, Goldstein, Holt, and Smith, 2005; Goldstein and Spiegelhalter, 1996).

A variety of means of accounting for prior attainment have been used. One has been looking at simple differences between scores in standardised tests (e.g. Weber, Martin, and Patterson, 2001). Another has been grouping pupils accord-

ing to their prior attainment scores, and comparing their subsequent performance with the average performance for pupils in the same prior attainment group (e.g. Department for Education and Skills, 2005). There are problems with both these approaches. For simple differences in test scores to be used, the tests themselves need to be specially designed so that a difference of  $x$  means the same if it comes from the difference between two high scores as if it comes from the difference between two low scores. These sorts of tests are typically designed for particular types of pupils, and they may not be well suited to other types of pupil. This makes them hard to use in large scale studies. In the U.K., standardised test scores are available to some extent from measures of progress that pupils make along the "key stages" of education. However, these scores are very coarse and are best thought of as ordered categories rather than continuous scores. Grouping pupils according to levels of prior attainment and then considering subsequent test scores is a technique that is used in work examining the value added by schools in the U.K. Although this technique can be used more widely than the technique of looking at simple differences, it does involve a loss of information: when the pupils are grouped according to prior attainment, the differences between pupils within prior attainment groups is lost to the analysis.

Another way of accounting for prior attainment is to include the prior attainment variable as a regressor in a model for current attainment. When this is done, there is no need for the prior and current tests to have been designed to have a known relationship (as when the simple difference is used), and because there is no grouping of the prior test results, the loss of information that occurs upon grouping is not an issue. However, although this method of accounting for prior performance may appear to be preferable, it does bring its own difficulties. This is because the prior attainment variable can be considered to be endogenous. That is, the influences that brought about the prior attainment score will also (at least in part) be influencing the current attainment score, above and beyond their influence through the prior test. This can be demonstrated by looking at a simple multilevel model for current attainment which includes a prior attainment variable as a regressor:

$$y_{ij} = \alpha + \beta_1 x_{ij} + \delta_i + \epsilon_{ij} \quad (1.1)$$

where  $y_{ij}$  is the current test score for pupil  $j$  in school  $i$ ,  $x_{ij}$  is the prior test score for the same pupil/school combination,  $\delta_i$  is the effect of school  $i$ ,  $\epsilon_{ij}$  is the random error associated with pupil  $j$  in school  $i$ ,  $\alpha$  is the intercept of the model and  $\beta_1$  is the slope associated with the influence the prior attainment has on the current test score. It might also be considered appropriate to allow the slope to vary according to school. In this case the  $\beta_1$  parameter would gain the subscript  $i$ .

Let us now consider the influences that have brought about the  $x_{ij}$ . There will naturally be an effect due to the pupil (who can be identified by the same  $i, j$  labelling as above), which we will call  $\epsilon'_{ij}$ . There will also be an effect due to the school that the pupil was in at the time of the prior test that we will call  $\delta'_i$ . We thus have

$$x_{ij} = \alpha' + \delta'_i + \epsilon'_{ij} \quad (1.2)$$

where  $\alpha'$  is the mean of the  $x_{ij}$ .

The  $\epsilon_{ij}$  in the model for  $y_{ij}$  and the  $\epsilon'_{ij}$  in the model for  $x_{ij}$  cannot reasonably be assumed to be independent of each other as they both relate to the effect of the individual pupil on the test scores, and could be considered to represent the ability of the pupil in whatever the tests are assessing. This means that the  $x_{ij}$  regressor in the model for  $y_{ij}$  is not independent of the random part of that model, and thus can be considered to be endogenous.

It is possible that the source of endogeneity could additionally be via the school effects. If the prior test scores were obtained in the same schools as the current test scores then the subscript  $i$  from the model for  $x_{ij}$  is the same as the subscript  $i$  in the model for  $y_{ij}$ . We thus have the effects  $\delta_i$  and  $\delta'_i$  at the school level, and it is difficult to assume that the effect of the school on the prior test is independent of the effect of the same school on the current test. The school level is thus contributing to the endogeneity of the  $x_{ij}$  variable.

Where additional levels exist in the model (e.g. class, geographical/administrative structures) then additional sources of endogeneity may exist.

### 2.3 Allowing for endogeneity in the multilevel model

The standard multilevel model, such as (1.1) is not an appropriate model when an endogenous regressor is present. If estimation of this misspecified model takes place using the commonly used multilevel modelling algorithms (such as the Iterative Generalised Least Squares algorithm used by MLwiN (Rasbash, Steele, Browne, and Prosser, 2005), EM algorithm used by HLM (Raudenbush, Bryk, Cheong, and Congdon, 2001) and Newton-Raphson algorithm used by SAS PROC MIXED (SAS Institute, 1992), inconsistent parameter estimates may result (see Spencer and Fielding, 2002; Spencer, 2002; Spencer, 2003 for further discussion). This is because in the standard multilevel model, for which these algorithms are designed, all regressors are considered to be exogenous.

In the field of econometrics, the problem of endogeneity has often been dealt with by using an instrumental variable approach. This uses a set of “instruments” that are correlated with the original set of regressors (including any that are endogenous), but which is independent of the model disturbance. This set is

used alongside the original set of regressors to obtain parameter estimates that are consistent. See, for example, Bowden and Turkington (1984) for further information on instrumental variable methods. Spencer and Fielding (2000) and Spencer and Fielding (2002) have investigated the use of instrumental variable estimation with multilevel models.

As well as addressing the endogeneity issue with instrumental variable estimation, Spencer and Fielding (2002) also used a Bayesian approach to the estimation of the multilevel model, using a directed graph and the BUGS software (Spiegelhalter, Thomas, Best, and Gilks, 1995). The conclusion of their paper was that the approach offered by the directed graph could have some advantages over the instrumental variable method. Spencer (2002) used a combination of directed graphs (through the BUGS software) and classical multilevel modelling methods (through the MLwiN software) to model data on teaching styles.

## **2.4 Combining multilevel modelling and structural equation modelling**

The aim of structural equation modelling is to examine the relationships that are hypothesised to exist between variables. The variables are connected to each other in a model which may be represented with equations, but is also commonly expressed in graphical form. The connections in the model may be direct influences of variables on each other, but may also be made via unmeasured factors or latent variables that (it is hypothesised) underlie the mechanisms that have created the dataset. The parameters of the model are chosen so that the covariances between the variables, as estimated from the model, come as close as possible to the actual covariances between the variables, as calculated from the data. More background on structural equation modelling can be found in Dunn, Everitt, and Pickles (1993), Schumacker and Lomax (1996).

One of the main advantages of SEM is that the models that are able to be defined can be more flexible than is allowed by traditional modelling techniques. However, until recently the fitting of models that allow for data structured in a hierarchical manner has proved problematic.

The availability of methods that combine multilevel modelling and SEM comes from two directions. From the multilevel modelling side, advances have occurred to allow SEMs to be fitted to hierarchically structured data (Browne, 2003), and the methodology is available via the multilevel modelling package MLwiN. From the SEM side, Bauer (2003), Bentler and Liang (2002), Curran (2003), Muthén (1994) have produced work that now enables hierarchically structured data to be used with SEMs. The methodologies developed are available via the SEM packages EQS (Bentler, 1995), Lisrel (Jöreskog and Sorbom, 1996) and M-Plus (Muthén and Muthén, 1998-2006).

### 3. Previous Analyses of the Teaching Styles Data

#### 3.1 Teaching styles data

In the 1970s, Bennett and his team sent questionnaires to 1500 primary school teachers in 871 schools in Lancashire and Cheshire, asking about the classroom techniques that they used. A principal components analysis and cluster analysis were undertaken to identify the different teaching styles that were used. The final classification that was used split the teachers into three groups: those that used formal teaching methods, those that used informal methods and those that used a mixture of formal and informal methods. To further investigate how pupils responded to these three styles, 12 teachers from each of the styles were chosen, and a range of attainment and personality tests were administered to the children in these teachers' classes at the beginning (September) and end (June) of the academic year. The teachers chosen were those that used teaching methods typical of the formal, mixed and informal styles. Full details of the study can be found in Bennett (1976).

#### 3.2 Results from Bennett (1976)

Bennett (1976) used the following model to investigate pupil progress in three subject areas: Reading, Mathematics and English:

$$y_{ij} = \alpha_{k(i)} + \beta_1(x_{ij} - \bar{x}) + \epsilon_{ij} \quad (3.1)$$

where  $y_{ij}$  is the attainment in the subject area in the June test for pupil  $j$  taught by teacher  $i$ ,  $x_{ij}$  is the attainment in the same subject area in the previous September (with  $i$  and  $j$  as before and  $\bar{x}$  the overall mean of the  $x_{ij}$ ),  $\beta_1$  reflects the impact of the September test on the June test, and  $\epsilon_{ij}$  is pupil-specific random variation. The  $\alpha_{k(i)}$  is the fixed effect of the teaching style used by teacher  $i$ , with  $k(i)$  referring to either formal, mixed or informal.

This model used by Bennett (1976) does not tackle the issue of the data having a hierarchical structure, nor the issue of endogeneity. Estimates obtained by Bennett are shown in Table 1. Note that Bennett (1976) does not report standard errors.

Table 1: Effects of teaching styles by subject from Bennett (1976)

	Formal style	Mixed style	Informal style
Reading	106.19	106.64	104.36
Mathematics	105.35	101.72	102.75
English	108.18	106.29	105.46

We see that for Reading, the informal teaching style is estimated to have the worst effect on pupil progress, whereas the formal and mixed methods are very similar. Mathematics and English both show the same pattern of an advantage for the formal teaching style.

### 3.3 Results from Aitkin, Anderson and Hinde (1981)

Aitkin, Anderson and Hinde (1981) carried out a reanalysis of Bennett’s work on pupil progress, taking into account the fact that the data comes from a hierarchical structure. The model they use (with a simple change in notation to facilitate comparison with Bennett’s model in section 3.2) is

$$y_{ij} = \alpha_{k(i)}(x_{ij} - \bar{x}) + \delta_i + \epsilon_{ij} \quad (3.2)$$

where the notation is as in section 3.2 with the additional  $\delta_i$  representing the random effect of teacher  $i$ .

Although the model used by Aitkin, Anderson and Hinde (1981) does tackle the issue of the data having a hierarchical structure, it does not tackle the issue of endogeneity. This variance components model was estimated using an EM algorithm and the estimates produced are shown in Table 2. No standard errors are given as the EM algorithm used to obtain estimates did not give them.

Table 2: Effects of teaching styles by subject from Aitkin, Anderson and Hinde (1981)

	Formal style	Mixed style	Informal style
Reading	105.0	103.6	106.0
Mathematics	103.6	99.8	103.5
English	107.4	103.4	105.8

Although Aitkin, Anderson and Hinde (1981) used slightly different definitions for their teaching styles, we now see that the informal style is marginally preferable for Reading. In all three subject areas, the mixed style appears worst. However, Aitkin, Anderson and Hinde (1981) report that no differences are “statistically significant”.

### 3.4 Results from Spencer (2002)

Recognising that neither the approach of Bennett (1976) nor Aitkin, Anderson and Hinde (1981) had tackled both of the issues of hierarchically structured data and endogeneity, Spencer (2002) examined the methods used in Spencer and



Fielding (2002). In this latter paper, two approaches had been taken to overcome the problem of endogeneity in a multilevel model. One of the approaches combined instrumental variable techniques with traditional multilevel modelling techniques, and the other approach had used Bayesian methods through the BUGS software.

In both cases, the model used by Spencer (2002) is the same as that used by Aitkin, Anderson and Hinde (1981) with the addition of an equation for the centred  $x_{ij}$ .

$$\begin{aligned} y_{ij} &= \alpha_{k(i)} + \beta_1(x_{ij} - \bar{x}) + \delta_i + \epsilon_{ij} \\ (x_{ij} - \bar{x}) &= \delta'_i + \delta'_{ij} \end{aligned} \quad (3.3)$$

When analysing the teaching styles data, Spencer (2002) found that it was not possible to use the instrumental variable approach due to the lack of suitable instruments. When using BUGS, Spencer (2002) found that some of the model parameters suffered from problems of non-convergence of the estimates coming from the Gibbs sampler. To overcome this, Spencer (2002) combined the BUGS approach with traditional multilevel modelling. The parameter estimate associated with the endogenous prior test score was one that had achieved convergence in the BUGS analysis, and its value was recorded. A traditional multilevel analysis was then carried out using the MLwiN software, with the parameter of the endogenous regressor constrained to this recorded value. The resulting parameter estimates from MLwiN thus respect the hierarchical nature of the data and also allow for the endogeneity in the model.

The estimates produced by the combined approach of Spencer (2002) are shown in Table 3. (with estimated standard errors in brackets).

Table 3: Effects of teaching styles by subject from Spencer (2002)

	Formal style	Mixed style	Informal style
Reading	105.3 (1.387)	106.1 (1.397)	104.7 (1.341)
Mathematics	103.8 (1.191)	102.7 (1.210)	103.2 (1.160)
English	107.7 (0.936)	106.6 (0.949)	105.6 (0.910)

Here we see that within each subject area, the rank order of the point estimates are the same as those produced by Bennett (1976). However, the standard errors are such that 95% confidence intervals overlap enormously, leading to the conclusion that there is insufficient evidence to claim differences between the results obtained by the different teaching methods.

## 4. Analysing the Teaching Styles Data in a Multivariate Framework with Multilevel SEMs

### 4.1 A multivariate framework

The work by Bennett (1976), Aitkin, Anderson and Hinde (1981) and Spencer (2002) modelled the Reading, Mathematics and English scores, but all constructed a separate model for each subject area. Using the model used by Aitkin, Anderson and Hinde (1981) and Spencer (2002) which allowed for the fact that the data has a multilevel structure, we could instead consider a multivariate framework for modelling, as below:

$$\begin{aligned} y_{Rij} &= \alpha_{Rk(i)} + \beta_{R1}(X_{Rij} - \bar{X}_R) + \delta_{Ri} + \epsilon_{Rij} \\ y_{Mij} &= \alpha_{Mk(i)} + \beta_{M1}(X_{Mij} - \bar{X}_M) + \delta_{Ri} + \epsilon_{Rij} \\ y_{Eij} &= \alpha_{Ek(i)} + \beta_{E1}(X_{Eij} - \bar{X}_E) + \delta_{Ei} + \epsilon_{Eij} \end{aligned} \quad (4.1)$$

The addition of  $R$ ,  $M$  and  $E$  subscripts correspond to the models for Reading, Mathematics and English respectively. Other notation is as in section 2.4.

The  $\delta_{Ri}$ ,  $\delta_{Mi}$  and  $\delta_{Ei}$  are the effects of teacher  $i$  on the three subject areas. These effects are likely to be correlated with each other, and if we consider using a multivariate framework for the modelling, we are able to model these correlations. Similarly, the  $\epsilon_{Rij}$ ,  $\epsilon_{Mij}$  and  $\epsilon_{Eij}$  are the effects associated with pupil  $j$  of teacher  $i$ . These are also likely to be correlated with each other, and with a multivariate framework, these correlations can be modelled.

### 4.2 Endogeneity

Whilst fitting the multivariate model in section 4.1, we also want to allow for the endogeneity in the model. Consider the regressor ( $x_{Rij} = \bar{x}_R$ ) in the Reading part of the model. This is the mean-centred September test result obtained by pupil  $j$  of teacher  $i$ . We may consider a model for this as below:

$$(X_{Rij} - \bar{x}_R) = \delta'_{Ri} + \epsilon'_{Rij} \quad (4.2)$$

where  $\delta'_{Ri}$  is the random effect associated with the teacher level of the model and  $\epsilon'_{Rij}$  is the random effect associated with the pupil level. It may be the case that teacher  $i$  has not had contact with the pupils prior to the September test, and thus not have had an influence on  $x_{Rij}$ . However, the  $\delta'_{Ri}$  teacher level effect will also contain any influences due to higher levels not represented in the model, such as school or neighbourhood. It may thus still be regarded as potentially related to the  $\delta_{Ri}$  in model (4.1), which will also contain effects of higher levels. Additionally, the  $\epsilon'_{Rij}$  pupil level effect is highly likely to be related

to the  $\epsilon_{Rij}$  pupil level effect in model (4.1). We may thus consider  $(x_{Rij} - \bar{x}_R)$  as an endogenous regressor.

Similarly, we may consider  $(X_{Mij} - \bar{x}_M)$  and  $(X_{Eij} - \bar{x}_E)$  to be endogenous regressors with models as in equations (4.3) and (4.4).

$$(x_{Mij} - \bar{X}_M) = \delta'_{Mi} + \epsilon'_{Mij} \quad (4.3)$$

$$(x_{Eij} - \bar{X}_E) = \delta'_{Ei} + \epsilon'_{Eij} \quad (4.4)$$

### 4.3 Allowing for endogeneity

We tackle the problem of endogeneity by using actively modelling the endogenous relationship. We will consider two different approaches: a path model based on equation (4.1) for the June test scores and equations (4.2), (4.3) and (4.4) for the September test scores, and a factor model where the teacher and pupil effects in these equations are regarded as realisations of underlying latent factors.

The key to a successful analysis here is to combine the path model or factor model with an estimation procedure that will exploit the flexibilities of the model to the full. Here we look at the use of the multilevel modelling package MLwiN and consider both its Markov Chain Monte Carlo (MCMC) and Iterative Generalised Least Squares (IGLS) estimation methods. This package is one of the world-leaders in multilevel modelling software. We also look at the package EQS and the Bentler and Liang (2002) method for fitting a multilevel structural equation model. As with MLwiN, the package EQS is one of the most-used worldwide for structural equation modelling.

### 4.4 Path model

A path model for the Reading, Mathematics and English marks corresponds to equations (4.1), (4.2), (4.3) and (4.4). It is repeated here in equation (4.5) so that it can be seen in consolidated form.

$$\begin{aligned} Y_{Rij} &= \alpha_{Rk(i)} + \beta_{R1}(x_{Rij} - \bar{x}_R) + \delta_{Ri} + \epsilon_{Rij} \\ Y_{Mij} &= \alpha_{Mk(i)} + \beta_{M1}(x_{Mij} - \bar{x}_M) + \delta_{Ri} + \epsilon_{Rij} \\ Y_{Eij} &= \alpha_{Ek(i)} + \beta_{E1}(x_{Eij} - \bar{x}_E) + \delta_{Ei} + \epsilon_{Eij} \\ (x_{Rij} - \bar{x}_R) &= \delta'_{Ri} + \epsilon_{Rij}' \\ (x_{Mij} - \bar{x}_M) &= \delta'_{Mi} + \epsilon_{Mij}' \\ (x_{Eij} - \bar{x}_E) &= \delta'_{Ei} + \epsilon_{Eij}' \end{aligned} \quad (4.5)$$

In order to allow for the endogeneity, this system of six models can be fitted simultaneously, with the teacher effects  $\delta_{Ri}, \delta_{Mi}, \delta_{Ei}, \delta'_{Ri}, \delta'_{Mi}$  and  $\delta'_{Ei}$  allowed

to covary fully with each other. The pupil effects  $\epsilon_{Rij}, \epsilon_{Mij}, \epsilon_{Eij}, \epsilon'_{Rij}, \epsilon'_{Mij}$ , and  $\epsilon'_{Eij}$  are also allowed to covary fully with each other.

#### 4.5 Path model with MLwiN

The path model (4.5) can be fitted in MLwiN with estimation undertaken with MCMC methods. To examine how well the parameters have converged, a burn-in period of 5000 iterations followed by a run of 5000 updates was undertaken and plots of the final 1000 parameter estimates obtained. These plots can be seen in Figure 1. It can be seen that even after the relatively long period of updating, the parameter estimates for the teaching styles are not converging. Much longer runs show no better convergence properties. This method of analysis is therefore not pursued further.

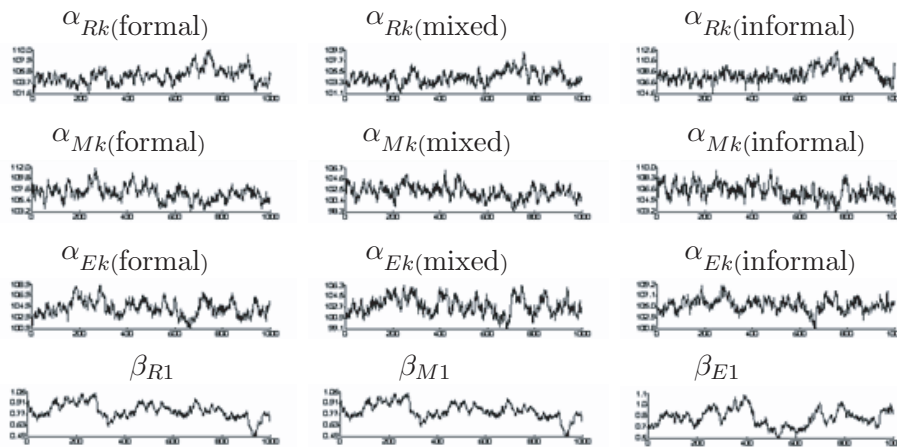


Figure 1: Updates from MLwiN MCMC estimation of path model

If we allow a misspecification of the model to remove the last three equalities in equation (4.5), we are then able to use MLwiN's IGLS estimation procedure. Because of the misspecification, we know that the parameter estimates produced will be inconsistent (see e.g. Spencer and Fielding, 2002). However, the dummy variables for the teaching styles exist at the upper level of the multilevel model. Being constant across the pupil level where the endogenous variable exists, they are not affected by the endogeneity, and will be estimated consistently. To demonstrate this effect, a series of simulations were carried out. The fifty simulated datasets were based on model (4.5) with arbitrary values used for the teaching style effects (but the same for each subject area for simplicity) and the effects of September tests on June tests. Table 4 shows the results of analysing these simulations using IGLS in MLwiN, and demonstrate the effect described above.

Table 4: Results of Simulations for MLwiN IGLS Estimation of Path Model

Mean Estimate from Simulations Effect	Target	(s.e.s in brackets)
Formal style on Reading	103.5	103.518 (0.0206)
Mixed style on Reading	108.5	108.485 (0.0258)
Informal style on Reading	98.4	98.378 (0.0216)
September Reading test on June test	0.5	0.785 (0.0353)
Formal style on Mathematics	103.5	103.505 (0.0191)
Mixed style on Mathematics	108.5	108.458 (0.0272)
Informal style on Mathematics	98.4	98.368 (0.0225)
September Mathematics test on June test	0.7	1.029 (0.0078)
Formal style on English	103.5	103.503 (0.0218)
Mixed style on English	108.5	108.470 (0.0208)
Informal style on English	98.4	98.381 (0.0228)
September English test on June test	0.75	1.096 (0.0052)

The misspecified model estimated with MLwiN and IGLS can therefore be used to obtain estimates of the teaching style effects for model (4.5). These are shown in Table 5. (with estimated standard errors in brackets).

Table 5: Effects of teaching styles by subject from MLwiN IGLS Estimation of Path Model

	Formal style	Mixed style	Informal style
Reading	105.0 (1.149)	106.8 (1.159)	104.5 (1.112)
Mathematics	104.3 (1.098)	102.2 (1.112)	102.8 (1.065)
English	108.0 (0.926)	106.3 (0.938)	105.4 (0.898)

The results obtained in this manner are very similar to those that can be seen in Table 3 produced in Spencer (2002) which also adjusted for endogeneity. There is a great deal of overlap in the 95% confidence intervals, leading again to the conclusion that there is insufficient evidence to claim differences between the results obtained by the different teaching methods.

#### 4.6 Path model with EQS

The path model in (4.5) has a large number of parameters. There are twelve fixed parameters and many more parameters due to variances and the covarying of the teacher effects  $\delta_{Ri}, \delta_{Mi}, \delta_{Ei}, \delta'_{Ri}, \delta'_{Mi}$  and  $\delta'_{Ei}$  and of the pupil effects  $\epsilon_{Rij}, \epsilon_{Mij}, \epsilon_{Eij}, \epsilon'_{Rij}, \epsilon'_{Mij}$  and  $\epsilon'_{Eij}$ . In fact, there are too many parameters to estimate via the SEM methods available in EQS. Alterations could be made to the model by restricting the number of parameters to be estimated. This may allow the SEM methods of EQS to be used. However, instead of doing this, we choose to explore the use of a factor model which requires fewer parameters to be estimated.

#### 4.7 Factor model

The factor model can be written as equation (4.6), below.

$$\begin{aligned}
 Y_{Rij} &= \alpha_{Rk(i)} + \beta_{R1}(x_{Rij} - \bar{x}_R) + \lambda_{\delta Ri}F_{\delta i} + \lambda_{\epsilon Rij}F_{\epsilon ij} \\
 Y_{Mij} &= \alpha_{Mk(i)} + \beta_{M1}(x_{Mij} - \bar{x}_M) + \lambda_{\delta Mi}F_{\delta i} + \lambda_{\epsilon Mij}F_{\epsilon ij} \\
 Y_{Eij} &= \alpha_{Ek(i)} + \beta_{E1}(x_{Eij} - \bar{x}_E) + \lambda_{\delta Ei}F_{\delta i} + \lambda_{\epsilon Eij}F_{\epsilon ij} \quad (4.6) \\
 (x_{Rij} - \bar{x}_R) &= \lambda'_{\delta Ri} + F'_{\delta i} \\
 (x_{Mij} - \bar{x}_M) &= \lambda'_{\delta Mi} + F'_{\delta i} \\
 (x_{Eij} - \bar{x}_E) &= \lambda'_{\delta Ei} + F'_{\delta i}
 \end{aligned}$$

In equation (4.6), the effect that the September tests ( $x_{Rij}, x_{Mij}, x_{Eij}$ ) have on the June tests ( $y_{Rij}, y_{Mij}, y_{Eij}$ ) are represented by the  $\beta_{R1}, \beta_{M1}$  and  $\beta_{E1}$ . The teacher level effects are represented by unknown latent factors  $F_{\delta}$  and  $F'_{\delta}$ , and the  $\lambda_{\delta}, \lambda'_{\delta}$  are related loadings. These two factors are allowed to covary. Similarly, the pupil level effects are represented by unknown latent factors  $F_{\epsilon}$  and  $F'_{\epsilon}$ , and the  $\lambda_{\epsilon}, \lambda'_{\epsilon}$  are related loadings. These factors are also allowed to covary. The teaching style effects are included in the factor model through the  $\alpha$ s with subscripts  $R, M$  and  $E$  for Reading, Mathematics and English.

#### 4.8 Factor model with MLwiN

The factor model of equation (4.6) can be fitted using MCMC estimation in MLwiN. As with the path model, an examination of the convergence of the parameter estimates must be undertaken. After a burn-in period of 5000 iterations, 5000 updates was undertaken and plots of the final 1000 parameter estimates obtained. These plots can be seen in Figure 2. It can be seen that even after a relatively long period of updating, parameter estimates for the teaching style

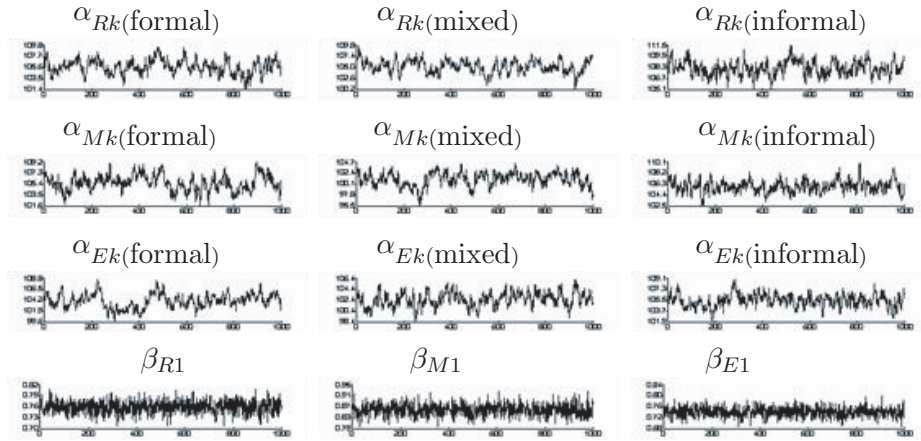


Figure 2: Updates from MLwiN MCMC estimation of factor model

Table 6: Results of Simulations for EQS Estimation of Factor Model

Mean Estimate from Simulations	Effect	Target (s.e.s in brackets)
Formal style on Reading	103.5	103.468 (0.0227)
Mixed style on Reading	108.5	108.360 (0.0288)
Informal style on Reading	98.4	98.448 (0.0196)
September Reading test on June test	0.5	0.506 (0.0059)
Formal style on Mathematics	103.5	103.463 (0.0217)
Mixed style on Mathematics	108.5	108.396 (0.0298)
Informal style on Mathematics	98.4	98.371 (0.0205)
September Mathematics test on June test	0.7	0.699 (0.0069)
Formal style on English	103.5	103.469 (0.0231)
Mixed style on English	108.5	108.399 (0.0276)
Informal style on English	98.4	98.385 (0.0210)
September English test on June test	0.75	0.739 (0.0058)

effects are not converging, although the story appears to be better for the effects of the September tests on the June tests. Much longer runs show no better convergence properties. This method of analysis is therefore not pursued further.

#### 4.9 Factor model with EQS

The factor model of equation (4.6) can be fitted successfully using EQS. Before discussing the data on teaching styles, we first undertake a simulation exercise to demonstrate the success of the method. The same fifty simulated datasets were used as discussed in section 4.5. Table 6 shows the results of analysing these simulations using EQS, and demonstrate the success of the method.

EQS can therefore be used to obtain estimates of the parameters for the model represented by equation (4.6). These are shown in Table 7. (with estimated standard errors in brackets). Note that the standard errors for each subject area are indeed the same across the teaching styles.

Table 7: Effects of teaching styles by subject from EQS Estimation of Path Model

	Formal style	Mixed style	Informal style
Reading	105.0 (0.637)	106.8 (0.637)	104.5 (0.637)
Mathematics	104.0 (0.663)	102.2 (0.663)	103.0 (0.663)
English	108.0 (0.534)	106.2 (0.534)	105.4 (0.534)

The point estimates are almost the same as those produced by MLwiN and IGLS for the path model (see Table 5). However, the standard errors are much smaller, and hence the confidence intervals narrower. There is still a good degree of overlap, but much less than before. Indeed now for English, the formal teaching style's interval does not overlap with that for the informal style.

#### 5. Conclusions

In this paper we have extended the work of previous authors by considering a multivariate model for modelling teaching styles data. We have examined two types of model: a path model and a factor model, and three methods of estimation: MCMC (through MLwiN), IGLS (also through MLwiN) and the Bentler and Liang (2002) method for fitting a multilevel structural equation model (through EQS). For a path model with an intentional misspecification, the IGLS estimation method has been successful in obtaining estimates of the effects of teaching styles. However, this success must be tempered by noting that the method does not deal with the issue of endogeneity in its entirety: the effects of teaching styles are only estimated consistently because they exist at the upper level of the multilevel model. The MCMC method as implemented in MLwiN was not able to obtain converging estimates for the path model. EQS could not fit the path model without needing to place restrictions on the number of



parameters to be estimated. For the path model, the MCMC method of MLwiN still have convergence problems. EQS however is successful in fitting the model. Indeed, this analysis is the most successful of all the analyses, past and present, as it obtains parameter estimates that have smaller standard errors than other methods.

## References

- Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A* **144**, 419-461.
- Bauer, D. J. (2003). Estimating multilevel linear models and structural equation models. *Journal of Educational and Behavioral Statistics* **28**, 135-167.
- Bennett, N. (1976). *Teaching Styles and Pupil Progress*. Open Books.
- Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Multivariate Software, Inc.
- Bentler, P. M. and Liang, J. (2002). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In *Multilevel Modeling: Methodological Advances, Issues and Applications* (Edited by N. Duan, S. Reise), 53-70. Lawrence Erlbaum Associates.
- Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T., and Smith, P. C. (2005). Performance indicators: good, bad and ugly. *Journal of the Royal Statistical Society, Series A* **168**, 1-27.
- Bowden, R. J. and Turkington, D. A. (1984). *Instrumental Variables*. Cambridge University Press.
- Browne, W. J. (2003). *MCMC Estimation in MLwiN, Version 2.0*. Institute of Education.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research* **38**, 529-569.
- Department for Education and Skills (2005). *National Curriculum Assessments at Key Stage 2 and Key Stage 1 to Key Stage 2 Value Added Measures in England, 2004/2005 (Revised)*. Department for Education and Skills.
- Dunn, G., Everitt, B., and Pickles, A. (1993). *Modelling Covariances and Latent Variables using EQS*. Chapman and Hall.
- Goldstein, H. (2003). *Multilevel Statistical Models*. Arnold.
- Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A* **159**, 385-409.
- Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Laurence Erlbaum Associates.

- Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Scientific Software International.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research* **22**, 376-398.
- Muthén, L. K. and Muthén, B. O. (1998-2006). *Mplus User's Guide. Fourth Edition*. Muthén and Muthén.
- Rasbash, J., Steele, F., Browne, W., and Prosser, B. (2005). *A User's Guide to MLwiN Version 2.0*. University of Bristol.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications.
- Raudenbush, S., Bryk, A. Cheong, Y. F., and Congdon, R. (2001). *HLM 5, Hierarchical Linear and Nonlinear Modelling*. Scientific Software International.
- SAS Institute (1992). *SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements*. SAS Institute, Inc.
- Schumacker, R. E. and Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Laurence Erlbaum Associates.
- Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications.
- Spencer, N. H. (2002). Combining modelling strategies to analyse teaching styles data. *Quality and Quantity* **36**, 113-127.
- Spencer, N. H. (2003). Consistent parameter estimation for lagged multilevel models. *Journal of Data Science* **1**, 123-147.
- Spencer, N. H. and Fielding, A. (2000). An instrumental variable consistent estimation procedure to overcome the problem of endogenous variables in multilevel models. *Multilevel Modelling Newsletter* **12**, 4-7.
- Spencer, N. H. and Fielding, A. (2002). A comparison of modelling strategies for value-added analyses of educational data. *Computational Statistics* **17**, 103-116.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1995). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.50*. MRC Biostatistics Unit.
- Weber, K., Martin, M. M., and Patterson, B. R. (2001). Teacher behaviour, student interest and affective learning: putting theory to practice. *Journal of Applied Communication Research* **29**, 71-90.

Received September 16, 2006; accepted December 9, 2006.

Neil H. Spencer  
Department of Marketing and Enterprise  
Business School  
University of Hertfordshire, de Havilland campus  
Hatfield, Hertfordshire, AL10 9AB, U.K.  
N.H.Spencer@herts.ac.uk