Chapter 13
Hierarchical Linear Models for Research on Professional Learning: Relevance and Implications

Eva Kyndt and Patrick Onghena

Abstract The goal of this book chapter is twofold. Because research on professional learning using Hierarchical Linear Modelling (HLM) is scarce, the first goal of this book chapter is to familiarise the readers with HLM. The opportunities, assumptions, and limitations of these techniques will be discussed and illustrated with an authentic dataset. Secondly, this chapter will focus on the relevance and implications of HLM for research in the field of professional learning and training: Why and when should this method be adopted within research on professional learning? Which conditions should be fulfilled? What are the advantages of this technique in general and which advantages are specifically relevant for the proposed field of research?

This chapter will provide a basic introduction into HLM. HLM, also known as multilevel modelling, is a statistical technique that takes the nested structure of the data into account. First, HLM will be introduced from a conceptual point of view and discussed why and when it could be applied. Subsequently, the different steps in HLM analysis will be presented and illustrated with an authentic dataset of a study investigating employees' learning intentions. The following section will focus on prior research in the field of professional learning that applied HLM. Finally, the main conclusions and future research perspectives will be provided.

Keywords Hierarchical linear modelling • Multilevel analysis • Professional learning

E. Kyndt (✉)
Centre for Research on Professional Learning and Development, Corporate Training
and Lifelong Learning, University of Leuven, Leuven, Belgium
e-mail: Eva.Kyndt@ppw.kuleuven.be

P. Onghena
Methodology of Educational Sciences Research Group, University of Leuven,
Leuven, Belgium

13.1   Introduction

The research field on professional learning has been growing considerably over the last decade (Billett 2008). While the field on professional learning has a long and rich tradition of applying qualitative research methods, nowadays more and more quantitative methods are being adopted to investigate the complex reality of professional learning. However the valid and relevant application of quantitative methods in this field is not self-evident: Employees with their own personal characteristics are embedded in different organisations that in turn have their own policies and characteristics. In addition, these organisations have to operate within a continuously changing society due to rapid technological changes and global competiveness (Kyndt et al. 2009). With any method a researcher uses, it is important the keep in mind that the data collection and analysis needs to be in accordance with the complex reality that is under investigation (Goldstein 2003). Therefore, it is our opinion that the field of professional learning would benefit from applying more advanced statistical techniques that are appropriate for analysing this complex reality.

The current chapter will focus on Hierarchical Linear Modelling (HLM), which is also known as multilevel modelling. The goal of this chapter is twofold. Because research on professional learning using HLM is scarce, the first goal of this chapter is to familiarise the readers with HLM. The opportunities, assumptions, and limitations of these techniques will be discussed. Secondly, this chapter will focus on the relevance and implications of HLM for research in the field of professional learning and training: Why and when should these methods be adopted in the research on professional learning? Which conditions should be fulfilled? What are the advantages of these techniques in general and which advantages are specifically relevant for the proposed field of research? The general aim of this book chapter is to provide a basic introduction to HLM without using mathematical formulas, without making it overly complex while at the same time staying true to the complexity of the presented analysis.

In the first section of this chapter, HLM will be introduced from a conceptual point of view and discussed why and when it could be applied. Subsequently, the different steps in HLM analysis will be presented and illustrated with an authentic dataset of a study investigating employees' learning intentions. The third section will focus on prior research in the field of professional learning that applied HLM. Finally, the main conclusions of this chapter will be summarized.

13.2   Hierarchical Linear Modelling (HLM): Why and When?

In order to present HLM from a conceptual perspective, the "why and when HLM should be used" will be discussed. We will explained that when the data are nested HLM should be used. In general two types of nested data can be distinguished; individuals can be nested within different contexts, that is employees within

organisations, or multiple measurement moments can be nested within individuals in longitudinal research (Raudenbush and Bryk 2002). Subsequently, the different types of relationships and models that can be investigated will be discussed, followed by which conditions should be fulfilled to be able to execute the analysis.

There are two main arguments for using HLM instead of regular regression analysis. The first argument pertains to the fact that the collected data need to be analysed in such a way that it takes the structure of the data into account. In other words, if the individuals from whom the data is collected are employed within different organisations – thus are nested in different organisations – then the analysis needs to take this nested structure into account. The second argument for using HLM refers to the theoretical grounding of the empirical research study. If theory and therefore hypotheses include variables at different levels then an analysis appropriate to handle these different levels is needed. For example, if theory describes that firm policy (equal for all employees of the same organisation) influences the individual job satisfaction of employees, an analysis that is able to investigate the relationship of firm policy at the organisational level with job satisfaction at the individual level is needed.

Both arguments can be related to the fact that '*ecological fallacy*' should be avoided. Ecological fallacy refers to the fact that relations are assumed to hold for individuals, whereas in reality they are observed in groups (Luke 2004). For example, research could identify that males and females differ in the number of trainings in which they participated, while in fact it is possible the differences are situated at the level of the organisation (e.g., industry, type of profession) but that some organisations have more or less female/male employees. Empirical research has indicated that occupational segregation could explain gender differences (Oosterbeek 1996; Simpson and Stroh 2002). In addition, disregarding the nested structure will lead to a higher likelihood of Type I errors. In other words, statistical relations are found that might not really exist. To ensure the reliability and validity of the findings, it is important that techniques are used that are appropriate for identifying the unique impact of specific factors at various levels (Snijders and Bosker 1999; Van den Noortgate et al. 2005).

Data collected in cross-sectional research may be nested, such as in the examples above, with individuals nested in organisations, but a specific type of nesting always occurs when performing longitudinal research. In longitudinal research, repeated measurements are nested within individuals. These models are also known as growth curve models, because researchers applying this technique are often interested in the growth trajectory of individuals (Anumendem et al. 2013; De Fraine et al. 2007; Prinzie et al. 2005). For example, research on transfer of training is concerned with the retention period of what has been learned, therefore they often use multiple measurement moments (e.g., Gegenfurtner 2013). Analysing growth curve models informs the researcher on the pattern or growth trajectory of the individuals.

The combination of both types of nesting is of course also possible; collecting longitudinal data with multiple measurement moments from individuals that are nested within different organisations. In this case an additional level is introduced in the model, resulting in a three-level model. Building further on the example of
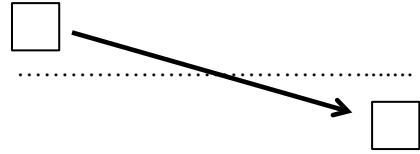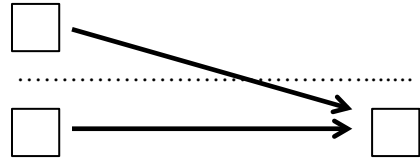
Fig. 13.1 Level-1 relationship
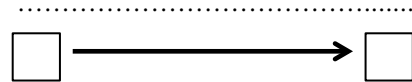
transfer of training; it could be that the collected longitudinal data stem from employees coming from different organisations. Another three-level model within the context of professional learning is for example an investigation of individuals nested within organisations that are in turn branches of larger (international) companies. These examples show that HLM offers a broad range of possibilities; this chapter will however limit itself to models with two levels, for more information of HLM including three or more levels we can refer the reader to the work of Goldstein (2003) or Raudenbush and Bryk (2002). In the second part of this chapter HLM analysis will be illustrated with a cross-sectional dataset involving two levels, for a good example of HLM analysis involving longitudinal data, the reader is referred to Van den Noortgate and Onghena (2006).

13.2.1 Types of Relationships

HLM analysis including two levels can involve different types of relationships. Even when no level-2 predictors are included in the model, because no characteristics of the organisation were collected, HLM is appropriate for nested data. Level-2 predictors are predictors that are identical for all the individuals within each organisation. Level-2 predictors can be (objectively) measured organisational characteristics, concern information retrieved from the human resource department of the company such as training budget or an aggregate of individual perceptions. However when aggregating individual perceptions, individual variation is lost therefore one can consider including these individual perceptions as a level-1 predictor. Figure 13.1 presents a type of relationship that includes one outcome variable and one predictor at the first level in a schematic way (Pustjens et al. 2004). The squares in the figure represent manifest variables, the arrows represent the relation between the variables and the dotted line separates the two levels in the model. The main difference between an HLM analysis of this type of relationship and regression analysis is that initial overall differences in the outcome variable between organisations are taken into account; this is not the case when using regression analysis. The article of Kyndt et al. (2013) is a published example of a research study within the field of professional learning investigating this type of relationship with HLM. They investigated the relationship between different predictors at the individual level (e.g., self-directedness, perceived support, etc.) and the learning intention of low-qualified employees while taking initial differences in learning intentions of employees belonging to different organisations into account (Kyndt et al. 2013).

When level-2 predictors have been measured, it can be investigated whether these organisational characteristics explain the variation observed at the individual level. For example, it could be investigated whether the company budget for training

Fig. 13.2 Cross-level relationship

Fig. 13.3 Cross-level relationship including level-1 predictor

**Fig. 13.4** Cross-level interaction
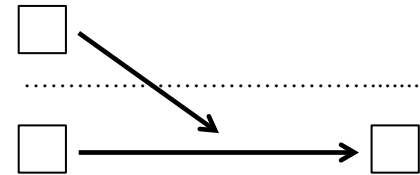
explains individual differences in participation in training. This type of relationship is called a cross-level relationship. This can be investigated apart or together with predictors at the first level. For example, company budget could be one of the predictors of training participation next to an individual's self-efficacy (e.g., Maurer et al. 2003). Figures 13.2 and 13.3 represent these relationships (Pustjens et al. 2004).

Finally, HLM allows us to investigate whether an interaction between characteristics at the level of the organisation and characteristics at the level of the individual explain the variation in the outcome variable at the individual level. For example, the interaction between the industry in which the organisation is active and the employee's level of education could predict the career development of employees. This relationship is known as a cross-level interaction effect and is represented in Fig. 13.4 (Pustjens et al. 2004). All figures in this section represent the relationship between manifest variables, however HLM can also be applied to latent variables (see e.g., D'Haenens et al. 2012; Muthén 1994). It is however important to point out that the outcome variables in HLM analysis need to be located at the lowest level that is included in the model. HLM analysis does not allow the prediction of a level-2 outcome based on level-1 predictors. For example, HLM cannot be used to predict the profits of an organisation based on the variation in individuals' level of education.

13.2.2   Different Types of Models

Before the different types of models that can be investigated with HLM are explained, the notion of fixed and random coefficients (intercept and regression coefficients determining the slope) needs to be introduced. Fixed coefficients are

Fig. 13.5 Fixed intercept
and slope
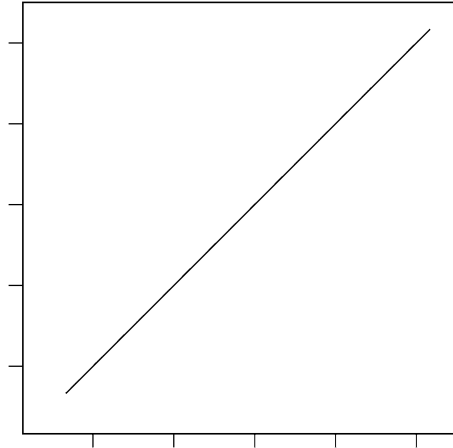
coefficients of which it is assumed that they hold for all cases in the data. When a
coefficient is said to be random, it means that its value can vary. However, random
is in itself a confusing term because it gives the impression that the coefficients can
take on any value, while this is in fact not the case (Field et al. 2012). An assumption
regarding these random coefficients is that they are normally distributed around the
average population coefficient.

Regression analysis operates from the assumption that all parameters are
fixed, that the score on the outcome variable for each individual can be pre-
dicted based on the same values of the intercept and regression coefficients.
Figure 13.5 presents a traditional regression line of a model with fixed inter-
cept and fixed slope.

Within HLM analysis the intercept, the slope or both can vary between organisa-
tions normally distributed around respectively the average intercept and slope that
holds for the population. HLM is able to assess the amount of variation at each level
(Raudenbush and Bryk 2002).

13.2.2.1 Random Intercept Model

Within a random intercept model the group effect is conceived as random – normally
distributed around the intercept of the population – rather than fixed (Raudenbush
and Bryk 2002). This type of model assumes homogeneity of the regression slopes
across the different groups or organisations included in the study, in other words it
assumes that the nature and strength of the relationship between the dependent and
independent variables is equal across the different organisations. This model does
however allow that initial differences in the overall level of the outcome variable
between organisations are taken into account. Figure 13.6 presents the regression
lines of the different organisations when a random intercept model is chosen.

Fig. 13.6 Random intercept
and fixed slope
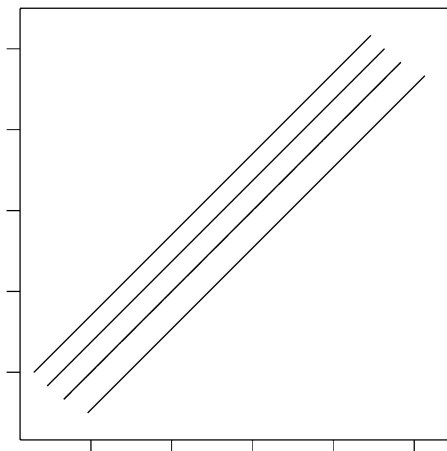
Fig. 13.7 Fixed intercept
and random slope
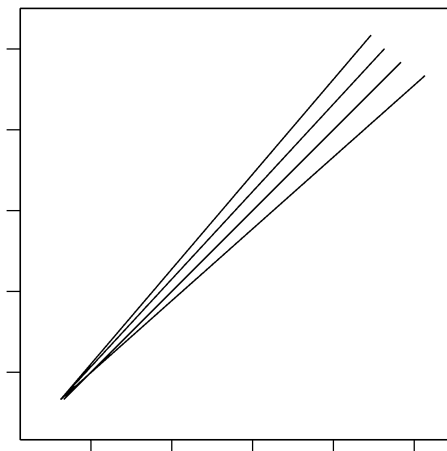
13.2.2.2   Random Slope Model

The random slope model represents the heterogeneity of the regression slopes. With
this model it is possible to investigate whether the relationship between the outcome
and predictor variables differs in nature and strength across organisations. In addi-
tion, this variance can possibly be explained by means of predictors situated at the
level of the organisation. This model does however assume that the intercept does
not vary across organisations. Figure 13.7 illustrates this model. In reality, this
model is rarely used because it is to be expected that variability in the nature of the
relationship (slopes) would normally create variability in the overall level of the
outcome variable (intercepts).

Fig. 13.8 Random intercept
and slope
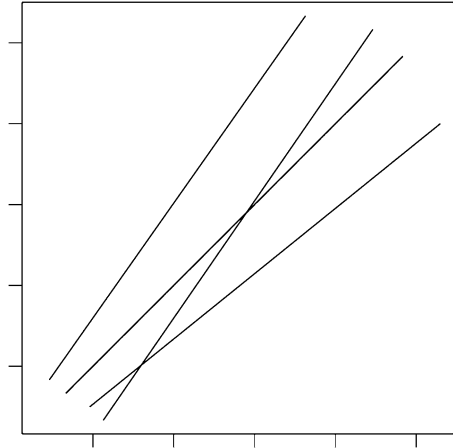
13.2.2.3 Random Intercept and Slope Model

Within this model both the intercepts and the slopes can vary across organisations. Both the initial differences in the overall level of the outcome variable and differences in the nature and strength of the relationship across organisations are being considered. Figure 13.8 represents this model.

13.2.3 Which Conditions Should Be Fulfilled?

13.2.3.1 Assumptions

Because HLM is an extension of regression analysis all assumptions of regression analysis apply to HLM. There is however one exception, in some cases the violation of independence can be resolved if this independence is caused by a level-2 variable, that is a variable at the level of the organisation. However, not all absence of independence can be explained by variables on the organisational level.

In addition, two supplementary assumptions are made that pertain to the random coefficients. In a random intercept model, it is assumed that the intercepts across the different organisations follow a normal distribution. In a random slopes model, this assumption also applies to the regression coefficients (Field 2009).

13.2.3.2 Sample Size

The complexity of HLM makes it difficult to formulate rules of thumb concerning the required sample size that can be applied to all datasets (Cools et al. 2008). However, some general guidelines can be offered together with the advice to execute

power analysis with the help of statistical software packages. In general, introducing another level in the analysis means that more parameters need to be estimated, and the more parameters that need to be estimated the larger the sample needs to be. It has been suggested that the number of groups are more important than the number of individuals in that group (de Leeuw and Kreft 1998). When interested in cross-level interactions the general guideline is that more than 20 groups/contexts (i.e., organisations) are needed and that each group size should not be "too small" (Field 2009). A more general but debateable rule of thumb is the 30/30 rule, meaning that at least 30 groups with 30 individuals are needed. However, as mentioned before the more complex the model, the larger the sample size needs to be and vice versa. For example, a simple HLM analysis (e.g., level-1 relationships with a random intercept) can be run with a dataset that is comparable to what would be needed to run a regression analysis.

The sample size of the collected data will also be important for the choice of the estimation method of the parameters, these estimation methods are discussed within the following section 'analysing and interpreting the data'.

13.3 Analysing and Interpreting the Data

Within this section the goal is to offer some guidelines for making decisions about the data and the steps that could be followed within the analyses, as well as the interpretation of the results. However, bear in mind that how the model is build and the decisions taken within this process need be theory driven. The procedure can differ depending on the subject and data at hand.

13.3.1 Illustration: Concept and Sample

Throughout this section the use of HLM will be illustrated with a dataset that was compiled to investigate the learning intentions of employees. The theory of reasoned action of Fishbein and Ajzen (1975) serves as the theoretical framework for defining a learning intention. Within this theory an individual's intention plays a central role in the decision making process of the individual (Baert et al. 2006). A learning intention can be defined as a readiness or even plan to undertake a concrete action in order to neutralise the experienced discrepancy, and to reach a desired situation by means of training and education (Kyndt et al. 2011). Within this investigation the role of self-efficacy, employability, self-directedness in career processes, time management, pay satisfaction, and perceived organisational support was investigated. These variables can all be situated at the level of the individual. The sector of the organisation (public versus private) was included as a level-2 predictor. A more complete theoretical background on learning intention and its predictors included in the dataset can be found in Kyndt et al. (Accepted).

Table 13.1 Educational level participants

Educational level	Frequency	Percentage	Cumulative percentage
No degree	48	3.9	3.9
Elementary school	28	2.3	6.2
Lower secondary education	124	10.2	16.4
Special needs education	28	2.3	18.7
Vocational secondary education	144	11.8	30.5
Higher secondary education	440	36.2	66.7
Professional bachelor	194	15.9	82.6
Master	193	15.9	98.5
Other	18	1.5	100
Missing	26		

The sample consisted of 1,243 employees (55 % female). The majority of these participants (82.3 %) were employed within 21 different organisations; the remaining 29.3 % of the participants did not provide the name of their organisation. Almost half of the participants (48.42 %) was employed within the public sector, the other 51.58 % was active in the private sector. The majority of the employees had a fulltime tenured position (61.8 %), 13.28 % had a part-time tenured contract, 10.66 % were temporarily employed and the remaining 14.26 % indicated having an 'other' type of contract (e.g., independent, constitutional appointment, etc.). On average, employees were 41.88 years old ($SD = 11.91$) and had 13.54 years of seniority ($SD = 12.85$). Table 13.1 contains the information regarding the educational level of the participants.

For this illustration the analyses were performed with the *lmne* package of R. R is free software for statistical computing that can be downloaded from www.R-project. org (R Development Core Team 2012). The R code of this example can be found in Appendix. HLM analysis can also be performed with SPPS, SAS MIXED procedure, HLM 7.0 or MLwiN. Each package has it advantages and disadvantages, see Tabachnick and Fidell (2001) or Twisk (2006) for a comparison of software for HLM analysis.

For this illustration, we chose to present the outputs as given by R so that the readers would recognise these outputs when undertaken the analysis themselves. These outputs present more information than discussed in this introductory chapter therefore we have marked the values on which the interpretations are based. We will explain step by step how the final model was built. When performing the analysis in R, the first steps that need to be undertaken are setting a working directory, loading the data, and installing the necessary packages. All information regarding the structure of that data, reliability of the scales, descriptive statistics and correlations can be found in Kyndt et al. (Accepted). For the clarity of the illustration the demographic characteristics of the participants will not be included in the multilevel model. Therefore results in terms of the research topic should be interpreted with caution. Readers interested in this research topic are referred to the article (Kyndt et al. Accepted).

The following steps in the analysis process will be explained and illustrated:

- Choosing an estimation method
- Assessing model fit and comparing different models

- Checking the need for HLM
- Centring the data
- Random intercept model with fixed predictors
- Random intercept and slope model
- Calculating effect sizes
- Reporting the results

13.3.2 Choosing an Estimation Method

A first step in the analysis involves choosing an estimation method for the parameters. In general two different estimation methods are used: Full Information Maximum Likelihood (FIML) and Restricted Maximum Likelihood (REML).

Both methods have their advantages and disadvantages, as already mentioned the sample size is one of the things that plays a role when deciding which estimation method to use. When the sample size is rather small, for example when only the minimum requirements discussed in the section 'sample size' are met, REML is more suited because FIML can lead to a negative bias in the estimates, especially when the number of parameters increases (i.e., additional predictors and random slopes). However, FIML has a large advantage when building the model. Later on it will be explained how different multilevel models can be compared in terms of how well they fit the data. When using FIML, it is possible to compare the fit of both the regression coefficients and variance estimates, whereas REML only allows comparing variance estimates (Peugh 2010). In other words, when comparing multilevel models REML only allows a researcher to conclude which model explains the most variance in the data. When comparing multilevel models estimated using FIML, it is also possible to draw conclusions about the comparison of regression coefficients, for example it can be determined whether a random intercept shows a better fit than a fixed intercept. For a more in depth discussion of these methods, the reader is referred to the work of Peugh (2010), Goldstein (2003) and Raudenbush and Bryk (2002).

Because the sample size in our illustration is sufficiently large ($n = 1,243$) and the interest lies in comparing the coefficients of different HML models, the FIML estimation method will be used for this example.

13.3.3 Assessing Model Fit and Comparing Different Models

Comparing different models is an essential part of HLM analysis, based on the comparisons of these models the (final) model is built up in search of the best model possible given the data. The overall fit of a multilevel model is tested by means of the Chi-square likelihood ratio test. To calculate this Chi-square likelihood ratio test, the -2log-likelihood statistic can be used. When comparing different nested models, that is models that built further on each other like in our example, a smaller

−2log-likelihood is better; however this value cannot be interpreted as such because there are no cut-off values that indicate a good or bad fit. It can only be concluded that in comparison with another model, a certain model has a better or worse fit with the data (Field et al. 2012). A significant Chi-square statistic indicates that the model with the lowest −2log-likelihood has a significantly better fit than the model to which it has been compared. Typically a new model that includes all the parameters of the old model complemented with new parameters is compared to the old model. Each new parameter that is included in the model will lead to a decrease in the −2 log likelihood, however the goal is to reach an as good as possible model fit with an as simple as possible model. Therefore the statistical significance of this difference is tested.

13.3.4   Checking the Need for HLM

Two different dominant views are present within the literature on HLM concerning the fact whether or not one should check if there is a need for HLM. On the one hand there is the view that this should not be checked, because HLM should always be applied when the data are nested (e.g., Goldstein 2003; Raudenbush and Bryk 2002). On the other hand, researchers have argued that nested datasets do not automatically require HLM (Peugh 2010). In the latter view, if there is no variation in response variables across organisations (level-2 units), there is no need for HLM and regression analysis can be used to analyse the data. In favour of the first view it can be said that it is not wrong to use HLM for nested datasets even if the variation across organisations is limited or non-existing. The results of HLM will resemble those of the regression analysis, but will from a conceptual point of view reflect the model in a more appropriate way.

However, from a pragmatic point of view, it can be argued that it is easier for researchers and readers with a limited amount of statistical knowledge to conduct and interpret regression analysis in comparison with HLM. Therefore also the second view has its merits. These researchers check the need for HLM by calculating the intraclass correlation (ICC) and the design effect statistics (Peugh 2010) based on the variance components of a random intercept model that does not include predictors (null model). The ICC reflects the proportion of variance of the dependent variable that can be explained by the mean scores of that same dependent variable across the organisations. A large ICC indicates that a large proportion of the variation in the dependent variable occurs at the level of the organisation and that the assumption of independence of regression analysis is violated. When the ICC is large, this means that differences are occurring at the level of the organisation rather than at the level of the individual. For example, when investigating the difference in participation in work-related learning amongst employees, a large ICC would indicate that differences in participation could be attributed to the differences at the organisational level (e.g., training budget) rather than differences at the individual level (e.g., age).

Calculating the ICC

Step 1: Calculate the null model.
Step 2: ICC = variance intercept/(variance intercept + variance residual)

The design effect is calculated based on the ICC score and the average number of employees within the organisation. Note that this average number of employees pertains to the average number of participants in the dataset that are nested in each organisation and not the size of the firm in terms of number of employees, which is often used as a predictor when investigating participation in work-related learning for example (Kyndt and Baert 2013). This design effect quantifies the effect of the violation of independence on standard error estimates; it quantifies the negative bias that results from nested data (Peugh 2010). According to Peugh (2010) a non-zero ICC combined with a design effect higher than 2, indicates the need for HLM (Muthén 1994; Peugh 2010).

Calculating the design effect

$$\text{Design effect} = 1 + \left(n_c - 1\right) \times \text{ICC}$$

n_c = average number of participants per organisation

Next, we will illustrate this approach with the data from our example. Output 13.1a shows the results of the random intercept null model predicting learning intention, subsequently it was calculated that the ICC equals .17 and design effect equals 9.03 (Output 13.1b). These values indicate the need for HLM because the ICC is larger than zero and the design effect is larger than 2 (Peugh 2010).

```
Random Intercept Model

Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC       BIC     logLik
  2661.007 2675.769 -1327.504

Random effects:
 Formula: ~1 | OrganisationID
         (Intercept)  Residual
StdDev:   0.4040179  0.8802076

Fixed effects: Intention ~ 1
               Value  Std.Error  DF  t-value  p-value
(Intercept) 3.069527 0.1008507  992 30.43634       0

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.88547944 -0.70480876  0.07414122  0.65850597  2.42413551

Number of Observations: 1013
Number of Groups: 21
```

Output 13.1a Random intercept model

```
Calculation variance:
        Variance   = SD²
        Variance intercept = .4040179² = .16323046
        Variance residual = .8802076² = .77476542

Calculation ICC:
        ICC = variance intercept/ (variance intercept + variance residual)
        ICC = .16323046 / (.16323046 + .77476542) = .17402044

Calculation average number employees per organisation:
        Nᴄ= 1013/21 = 48.24

Calculation Design effect:
        Design effect = 1 + (nᴄ -1) x ICC
        Design effect = 1 + (48.24-1) x .17
        Design effect = 9.0308
```

Output 13.1b Calculation of ICC and design effect

Alternatively when using FIML, the need for HLM can also be checked by calculating a null model with a fixed intercept (using maximum likelihood estimation) and a null model including a random intercept and subsequently test with the Chi-square likelihood ratio test which model fits the data best. However, only the fit of two models that are identical with the exception that in the first model the intercept is fixed and in the second model the intercept is random, can be compared. If the Chi-square likelihood ratio test is significant and the random intercept model has the smallest −2log-likelihood value, a HLM analysis should be applied to the data.

It is important to notice that for this comparison the same number of subjects should be included in both analyses. When a complete dataset is available this will not be a problem. However, when confronted with missing values a listwise deletion of missing values is the easiest way to achieve this equality. Output 13.2 shows the results of both models and their comparison. The Chi-square likelihood ratio test is significant and the −2log likelihood of the random intercept model is smaller than the −2log likelihood of the model with a fixed intercept. These results confirm the need for HLM. In other words, a sufficiently large proportion of the variance of employees' learning intentions can be situated at the level of the organisation. For the purpose of this book chapter both methods were illustrated, of course it is sufficient to execute one of these methods.

13.3.5   Centring the Data

In HLM analysis it is often useful to centre the predictor variables, in fact very few situations are suitable for not centring these variables (Peugh 2010). Centring involves rescaling the predictor variable by subtracting a mean score of the predictor from each individual score. This is done so that a value of zero can be interpreted meaningfully as the central tendency of the distribution (Field 2009; Peugh 2010). This is recommended for predictor variables within professional learning because a

```
Model 1: Fixed intercept model (disregarding the nested structure of the data)

Generalized least squares fit by maximum likelihood
  Model: Intention ~ 1
  Data: DataMultilevel
      AIC      BIC    logLik
  2763.264 2773.105 -1379.632

Coefficients:
              Value  Std.Error  t-value p-value
(Intercept) 3.190918 0.02969272 107.4647       0

Standardized residuals:
        Min           Q1          Med           Q3          Max
-2.531188515 -0.625586063  0.009614754  0.856549177  1.915217206

Residual standard error: 0.9445832
Degrees of freedom: 1013 total; 1012 residual


Model 2: Random Intercept Model

Linear mixed-effects model fit by maximum likelihood
  Data: DataMultilevel
      AIC      BIC    logLik
  2661.007 2675.769 -1327.504


Random effects:
 Formula: ~1 | OrganisationID
        (Intercept)   Residual
StdDev:   0.4040179  0.8802076

Fixed effects: Intention ~ 1
              Value Std.Error  DF  t-value p-value
(Intercept) 3.069527 0.1008507 992 30.43634       0

Standardized Within-Group Residuals:
       Min          Q1          Med          Q3          Max
-2.88547944 -0.70480876  0.07414122  0.65850597  2.42413551

Number of Observations: 1013
Number of Groups: 21

Comparison fixed and random intercept model:
InterceptOnly:          -2log Lik. = 2759.264 (df=2)
randomInterceptOnly:    -2log Lik. = 2655.007 (df=3)

Model                   df    AIC      BIC    logLik   Test   L.Ratio  p-value
interceptOnly           1   2 2763.264 2773.105 -1379.632
randomInterceptOnly     2   3 2661.007 2675.769 -1327.504 1 vs 2 104.2569  <.0001
```

Output 13.2 Fixed intercept and random intercept model

score of zero usually has no intrinsic substantive meaning or because a score of zero is not within the range of scores for which interesting conclusions can be derived (e.g., Likert scales from 1 to 5). For example, if one would be interested in the relationship between an individual's IQ and his monthly wage, centring would be recommended because an IQ of zero and wage of zero do not occur, therefore interpreting an intercept of zero is meaningless within the relation between IQ and monthly wage. In addition, centring can partly resolve multicollinearity between predictor variables and centred multilevel models are more stable (Field 2009).

Two forms of centring are common for HLM analysis: grand mean centring and group mean centring. *Grand mean centring* involves subtracting the sample mean

(the mean score of all individuals included in the sample) from the predictor score of each individual. With *group mean centring* the mean of the individuals in the group (or organisation) is subtracted from the predictor score of each individual in that group. In this case, the mean is calculated for each organisation separately and then subtracted from the predictor score belonging to the employee of that organisation. Grand mean centring does not change the model, meaning that is comparable across groups. Because group mean centring involves subtracting a (possibly) different mean for each group it is logical that the models cannot be compared as such across the different organisations. Therefore Peugh (2010) advises to use group mean centring when only level-1 predictors are included in the analysis, because when only investigating level-1 relationships it is only necessary to merely control for the nested structure of the data and group mean centring is appropriate. When interested in differences between organisations, it is necessary to be able to compare the model across the different organisations. Therefore when level-2 predictors (i.e., organisational characteristics) are included in the analysis Peugh (2010) recommends grand mean centring. Enders and Tofighi (2007) make four recommendations that are partially in line with Peugh's (2010) recommendations. Their recommendations start from the research questions of the empirical study at hand. They advice group mean centring if the primary interest focuses on the association between level-1 predictors (a). For example, group mean centring is appropriate when investigating the relation between an employee's motivation and his approach to learning at work. In this case one can say that it is merely controlled for the fact that individuals are nested within organisations. When the primary interest is on the level-2 predictors while controlling for level-1 predictors, Enders and Tofighi (2007) advice grand mean centring (b). A possible research question could for example focus on the relationship between the industries and employees' participation in formal learning activities. In their view, and in contrast to Peugh (2010) both types of centring are possible when looking at the differential influence of a variable at level-1 and level-2 (c). In other words, when investigating the relationship between both individual and organisational characteristics on individual outcomes, such as the influence of learning attitude and sector on employees' job satisfaction. Finally, group mean centring is preferable for examining cross-level interactions (Enders and Tofighi 2007), such as the interaction between employees' educational level and a firm's training policy (d).

The choice between group mean and grand mean centring only applies to level-1 predictors. Group mean centring cannot be applied to level-2 predictors because it is an inherent characteristic of these predictors to be equal for all individuals within the same organisation (Enders and Tofighi 2007). For level-2 predictors, a researcher can choose between the raw score or grand mean centring (Enders and Tofighi 2007). Because only predictors are centred, there is no need to centre the data before the need for HLM is determined because the need for HLM is determined based on the model that only includes the intercept.

The study that is used as an illustration in this chapter includes both level-1 and level-2 predictors, in line with the guidelines of Peugh (2010) the predictors self-efficacy, time management, perceived organisational support, self-directedness in

career processes, pay satisfaction and employability will be centred by means of the grand mean centring method. The sample mean of each variable is subtracted from the score of each subject.

13.3.6 Random Intercept Model with Fixed Predictors

HLM analyses involve several steps in which a model is built up. When the need for HLM is established, the analyses are continued starting from the empty random intercept model. In a first step fixed predictor variables are added to the null model. Like in ordinary least-squares regression analysis this can be done stepwise or full subsets of predictors can be used. The analysis can either be started by adding the variables that are hypothesized to make the most important contribution to the model, followed by the second most important variable, etc. Each time it should be tested if this results into a better model than before. In this example, the predictor self-directedness in career processes was added to the model. Output 13.3 presents these results. The model including self-directedness shows a better fit than the model containing no predictors. In addition self-directedness is a significant predictor of an employees' learning intention when initial organisational differences in terms of the intercept are taken into account.

```
Model 3: Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC      BIC    logLik
  2337.415 2357.097 -1164.707

Random effects:
 Formula: ~1 | OrganisationID
        (Intercept)   Residual
StdDev:   0.2742968 0.7521469

Fixed effects: Intention ~ Cdirectedness
                 Value  Std.Error   DF  t-value p-value
(Intercept)   3.0805839 0.07195713 992 42.81138       0
Cdirectedness 0.7329617 0.03734167 991 19.62852       0


Correlation:
              (Intr)
Cdirectedness 0.007

Standardized Within-Group Residuals:
       Min        Q1        Med        Q3        Max
-3.9880997 -0.6514354  0.1396961  0.6715072  2.6470453

Number of Observations: 1013
Number of Groups: 21

Comparison with random intercept model:
                    Model df     AIC      BIC    logLik     Test   L.Ratio  p-value
randomInterceptOnly     1  3 2661.007 2675.769 -1327.504
randomInterceptDirect   2  4 2337.415 2357.097 -1164.707 1 vs 2 325.5924  <.0001

randomInterceptDirect:   -2log Lik. = 2329.415 (df=4)
```

Output 13.3 Random intercept model with self-directedness as a fixed predictor

```
Model 4: Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC       BIC     logLik
  2265.702 2309.988 -1123.851

Random effects:
 Formula: ~1 | OrganisationID
         (Intercept)    Residual
StdDev:   0.2598758 0.7225583
Fixed effects: Intention ~ Cdirectedness + Ctime + Cefficacy + Cemployability +
Csupport + Cfinancial
                     Value  Std.Error  DF  t-value p-value
(Intercept)      3.0837485 0.06871690 986 44.87613  0.0000
Cdirectedness    0.5717715 0.04412589 986 12.95773  0.0000
Ctime            0.1426350 0.02709274 986  5.26470  0.0000
Cefficacy       -0.0268543 0.05188570 986 -0.51757  0.6049
Cemployability   0.1299607 0.03602337 986  3.60768  0.0003
Csupport         0.1608344 0.03882145 986  4.14293  0.0000
Cfinancial      -0.0060371 0.02908900 986 -0.20754  0.8356

Correlation:
              (Intr) Cdrctd Ctime  Cffccy Cmplyb Cspprt
Cdirectedness  0.023
Ctime          0.220 -0.225
Cefficacy     -0.051 -0.387 -0.088
Cemployability 0.013 -0.081 -0.011 -0.344
Csupport       0.007 -0.156 -0.164  0.005 -0.137
Cfinancial    -0.031  0.099 -0.010 -0.090  0.014 -0.451

Standardized Within-Group Residuals:
       Min          Q1        Med          Q3        Max
-3.34828578 -0.67519565  0.09474177  0.69929517  2.75349733

Number of Observations: 1013
Number of Groups: 21

Comparison with random intercept model with one predictor
                    Model df     AIC      BIC    logLik   Test  L.Ratio p-value
randomInterceptDirect   1  4 2337.415 2357.097 -1164.707
randomInterceptFUll     2  9 2265.702 2309.988 -1123.851 1 vs 2 81.71326  <.0001


randomInterceptFUll:      -2log Lik. = 2247.702 (df=9)
```

Output 13.4 Random intercept model including all level-1 predictors (fixed effects)

Besides this stepwise forward method, a full subset method can be chosen. In this example all our predictors were inserted at the same time into the model. Output 13.4 shows the results of the model containing all level-1 predictor variables included in this research study. Self-directedness in career processes, time management, pay satisfaction, employability, perceived organisational support, and self-efficacy are simultaneously included in the model to predict an employees' learning intention. The results show that directedness, time management, employability and perceived organisational support are significant positive predictors of an employees' learning intention. Self-efficacy and pay satisfaction are not significant.

Now it can be decided to continue with this 'full' model that comprises all the level-1 predictor variables or to remove the non-significant predictors. In this case, it can be tested whether a model only containing the significant predictors still yields a better fit in comparison with the previous model that showed an improved model fit. For our example this means that self-efficacy and pay satisfaction will be removed from the model. Output 13.5 shows that the model without these

```
Model 5: Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC       BIC    logLik
  2262.036 2296.481 -1124.018



Random effects:
 Formula: ~1 | OrganisationID
         (Intercept)  Residual
StdDev:   0.2607045 0.7226461

Fixed effects: Intention ~ Cdirectedness + Ctime + Cemployability + Csupport
                  Value  Std.Error  DF  t-value p-value
(Intercept)     3.0812752 0.06868172 988 44.86310   0e+00
Cdirectedness   0.5636445 0.04055550 988 13.89810   0e+00
Ctime           0.1412574 0.02695992 988  5.23953   0e+00
Cemployability  0.1233618 0.03378571 988  3.65130   3e-04
Csupport        0.1564820 0.03459019 988  4.52389   0e+00
 Correlation:
                (Intr) Cdrctd Ctime  Cmplyb
Cdirectedness    0.006
Ctime            0.015 -0.282
Cemployability  -0.005 -0.247 -0.044
Csupport        -0.010 -0.152 -0.193 -0.171

Standardized Within-Group Residuals:
       Min          Q1        Med          Q3        Max
-3.3777093 -0.6684023  0.1034987  0.6993091  2.7492216

Number of Observations: 1013
Number of Groups: 21

Comparing full model with model without non-significant predictors
                     Model df     AIC      BIC    logLik    Test  L.Ratio p-value
randomInterceptDirect   1  4  2337.415 2357.097 -1164.707
randomInterceptFull2    2  7  2262.037 2296.481 -1124.018 1 vs 2 81.37832  <.0001


randomInterceptFull2:     -2log Lik. = 2248.036 (df=7)
```

Output 13.5 Removing non-significant predictors

non-significant predictors still has an improved model fit in comparison with the model only containing self-directedness. Because Model 5 contains fewer predictors than Model 4, the −2log likelihood is slightly higher. However Model 5 is a simpler model than Model 4 and shows a fit that is not significantly better or worse than the model including the non-significant predictors. Especially, when the dataset is limited and additional parameters need to be included, we would give the advice to remove non-significant predictors in order to avoid an over-parameterisation of the model, which may lead to convergence problems.

13.3.7 Random Intercept and Slope Model

When it has been determined that the intercept is random, researchers usually built further on this model because the variability in the slopes would normally create variability in the overall level of the outcome variable (intercepts).

When it is expected that the relationship between the predictor variable and the outcome variable may vary between different organisations, a random slopes model can be tested. To determine whether the variance in the slope is significant, two models including the same parameters for which the only difference is that the parameters are fixed instead or random can be compared. After testing the random intercept and regression coefficients, level-2 variables that could explain the variation in the slopes and intercept can be introduced.

It is important that it makes sense from a conceptual point of view that the predictor variable of which the slope is set to random can actually vary between organisations. In other words it should be possible to hypothesize from the theoretical background that the relationship between a predictor variable and outcome variable can potentially vary between different organisations. In this example, it is hypothesized that the relationship between perceived organisational support and an employees' learning intention could vary across organisations. Output 13.6 indeed shows that including a random slope for the organisational support improves the model fit. Within this

```
Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC      BIC    logLik
  2259.797 2304.083 -1120.899

Random effects:
 Formula: ~Csupport | OrganisationID
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.24815049 (Intr)
Csupport    0.08474624 0.838
Residual    0.71999459

Fixed effects: Intention ~ Cdirectedness + Ctime + Cemployability + Csupport
                  Value   Std.Error  DF  t-value p-value
(Intercept)    3.0810959 0.06615086 988 46.57681  0e+00
Cdirectedness  0.5746528 0.04046762 988 14.20031  0e+00
Ctime          0.1418792 0.02686214 988  5.28175  0e+00
Cemployability 0.1245083 0.03363032 988  3.70226  2e-04
Csupport       0.1408532 0.04126539 988  3.41335  7e-04
 Correlation:
               (Intr) Cdrctd Ctime  Cmplyb
Cdirectedness   0.004
Ctime           0.014 -0.288
Cemployability -0.003 -0.249 -0.041
Csupport        0.389 -0.141 -0.159 -0.130

Standardized Within-Group Residuals:
       Min         Q1        Med        Q3        Max
-3.16300632 -0.69456599 0.09055579 0.70308260 2.73027846

Number of Observations: 1013
Number of Groups: 21

Comparing random intercept model with fixed effect to the random slope model
                    Model df    AIC      BIC    logLik   Test  L.Ratio p-value
randomInterceptFull2    1  7 2262.037 2296.481 -1124.018
addRandomSlope          2  9 2259.797 2304.083 -1120.899 1 vs 2 6.239108  0.0442

addRandomSlope:      -2log Lik. = 2241.80 (df=9)
```

Output 13.6 Random intercept and random slope model

random intercept and random slope model, perceived organisational support is a significant random predictor for an employee's learning intention. Self-directedness, employability and time management remain significant fixed predictors of an employee's learning intention. The relationship of these three latter variables to an employee's learning intention is assumed to be equal within every organisation, whereas the relationship between organisational support and learning intention varies between organisations. Based on this model however, the variance between organisations based on organisational characteristics or level-2 predictors cannot be explained. It can only be concluded that this relationship varies.

After testing if a random slope model yields a better fit, level-2 predictors can be added to the model. In our example, the only level-2 predictor that was included was the sector (public versus private) of the organisations. First sector is added as a fixed predictor. Output 13.7 demonstrates that including sector in the model resulted into a better fit. The public sector was given the code '0', while the private sector was

```
Model 7: Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC      BIC    logLik
  2255.958 2305.165 -1117.979

Random effects:
 Formula: ~Csupport | OrganisationID
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.22400908 (Intr)
Csupport    0.08862837 0.995
Residual    0.71938149

Fixed effects: Intention ~ Cdirectedness + Ctime + Cemployability + Csupport +
Sector
                   Value  Std.Error  DF  t-value p-value
(Intercept)     3.242574 0.08450946 988 38.36936  0.0000
Cdirectedness   0.580665 0.04026664 988 14.42049  0.0000
Ctime           0.142282 0.02679924 988  5.30917  0.0000
Cemployability  0.122211 0.03344320 988  3.65430  0.0003
Csupport        0.142633 0.04078178 988  3.49747  0.0005
Sector         -0.277918 0.10418827  19 -2.66746  0.0152
 Correlation:
               (Intr) Cdrctd Ctime  Cmplyb Cspprt
Cdirectedness   0.010
Ctime           0.007 -0.293
Cemployability -0.017 -0.248 -0.038
Csupport        0.458 -0.130 -0.162 -0.142
Sector         -0.690 -0.014  0.006  0.023 -0.137


Standardized Within-Group Residuals:
        Min         Q1        Med         Q3        Max
-3.06400439 -0.68286129 0.09259604 0.69382784 2.73321922
Number of Observations: 1013
Number of Groups: 21

Comparing random slope model to random slope model with level-2 predictor
                Model df      AIC      BIC    logLik   Test L.Ratio p-value
addRandomSlope      1  9 2259.797 2304.083 -1120.899
addRandomLevel2     2 10 2255.958 2305.164 -1117.979 1 vs 2 5.83954  0.0157

addRandomLevel2:   -2log Lik. = 2235.96 (df=10)
```

Output 13.7 Random slope model with level-2 predictor

coded '1'. The results show that employees in the public sector have higher learning intentions than those in the private sector.

Subsequently, a cross-level interaction effect between sector and organisational support is added to explore whether sector is able to explain the variation in the slope of organisational support. Output 13.8 shows that this cross-level interaction does not improve the model fit. Because the random slopes model including sector as a level-2 predictor (Model 7) was the last model that increased the model fit, this model is considered as the final model on which the conclusions are based.

```
Model 8: Linear mixed-effects model fit by maximum likelihood
 Data: DataMultilevel
       AIC       BIC    logLik
  2257.811 2311.939 -1117.906

Random effects:
 Formula: ~Csupport | OrganisationID
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.22232708 (Intr)
Csupport    0.08822628 0.996
Residual    0.71940551

Fixed effects: Intention ~ Cdirectedness + Ctime + Cemployability + Csupport +
Sector + Csupport:Sector
                  Value   Std.Error  DF  t-value p-value
(Intercept)     3.225260 0.09588197 987 33.63782  0.0000
Cdirectedness   0.580085 0.04032672 987 14.38463  0.0000
Ctime           0.142413 0.02681203 987  5.31153  0.0000
Cemployability  0.122434 0.03346191 987  3.65891  0.0003
Csupport        0.126164 0.05946629 987  2.12161  0.0341
Sector         -0.251678 0.12464290  19 -2.01919  0.0578
Csupport:Sector 0.029388 0.07780262 987  0.37772  0.7057

Correlation:
                (Intr) Cdrctd Ctime  Cmplyb Cspprt Sector
Cdirectedness    0.031
Ctime            0.002 -0.293
Cemployability  -0.024 -0.249 -0.038
Csupport         0.625 -0.055 -0.117 -0.111
Sector          -0.770 -0.038  0.010  0.029 -0.482
Csupport:Sector -0.480 -0.047  0.008  0.019 -0.728  0.554

Standardized Within-Group Residuals:
        Min         Q1        Med         Q3        Max
-3.03787659 -0.69952968 0.09729221 0.68859241 2.72806969
Number of Observations: 1013
Number of Groups: 21

Comparing random slope model to random slope model with level-2 predictor
                Model df    AIC      BIC    logLik    Test  L.Ratio p-value
addRandomLevel2     1 10 2255.958 2305.164 -1117.979
CrossLevel:         2 11 2257.811 2311.939 -1117.906 1 vs 2 0.1466011  0.7018

CrossLevel:          -2log Lik. = 2235.81 (df=11)
```

Output 13.8 Cross-level interaction

13.3.8 Calculating Effect Sizes

The calculation of effect sizes in HLM analysis is a debated topic and currently no consensus exists with regard to which effect sizes are the most appropriate. In addition, it can be noticed that these effect sizes are actually very rarely reported in published articles. Probably due to the lack of consensus and the difficulty in calculating these effect sizes. However, within the framework of this chapter, it is important to provide the reader with a basic knowledge of effect sizes in HLM.

Below the basic effect sizes that are generally accepted will be presented (Raudenbush and Bryk 2002). Overall two types of effect sizes are distinguished: effect sizes that represent the variance in the outcome variable that is explained by all predictors, and the effect sizes that represent the effect of a specific variable on the outcome variable. When interested in the variance that is explained by all predictor variables, the predicted score for each individual in the dataset can be calculated. In most software programs the option to save these predicted values can be chosen when calculating the final model. Subsequently, the correlation coefficient between the predicted and observed scores of the outcome variable needs to be calculated and squared. This squared correlation coefficient is known as the pseudo-R^2 (Peugh 2010; Raudenbush and Bryk 2002).

Calculating the effect size of each variable separately is more complicated in comparison with the pseudo-R^2. This effect size is called the proportional reduction in variance (i.e., reduction in residual variance). However, to calculate this effect size for each variable separately, each variable would have to be added separately to the model because this effect size simply informs the researcher about how large the proportion of residual variance reduction is in comparison with the prior model that did not include the predictor for which the effect size is being determined (Peugh 2010; Roberts and Monaco 2006). The proportional reduction in variance (PRV) can be calculated as follows:

$$PRV = \left(var_{\text{prior model}} - var_{\text{model with predictor}}\right) / var_{\text{prior model}}.$$
$$var = \text{residual variance model}$$

However, some caution is needed when interpreting this effect size because the results can be counterintuitive. For example, this *PRV* can be larger than the overall pseudo-R^2, however these two effect sizes are not comparable and should not be interpreted in relation to each other (Peugh 2010). In addition, it is possible to obtain negative values for *PRV* when for instance level-2 predictors are included in the model (Roberts and Monaco 2006). Roberts and Monaco (2006) present of an overview of adjustments to the calculation of this effect size proposed in order to solve this problem.

Besides calculating the above-described effect sizes there are other ways to determine which variables contribute the most or the least to the model. By standardizing the variables included in the research, standardized coefficients are

calculated that can be compared. In addition the ICC discussed above, can be interpreted as an effect size of the random effects (Roberts and Monaco 2006). Readers who wish to know more about the effect sizes in HLM are referred to Hedges (2007, 2011).

13.3.9 Reporting the Results

Multilevel models can take on many different forms, making it difficult to offer one template that can be applied to all types. As was illustrated throughout this chapter, performing HLM analyses includes multiple steps in which a model is built up. It is advisable to report on all of these stages (Field 2009) as well as the software used to calculate the model. In addition, it is important to report the estimation method and the type of centring that were adopted.

For each model the F-statistics or t-statistics, degrees of freedom, and p-values should be reported. In addition, the variance components (random effects) and $-2\log$ likelihood should be presented. The output in R presents the standard deviations of the random coefficients, these need to squared in order to obtain the variance components. For the model comparison statements need to be underpinned by reporting the Chi-square statistic, degrees of freedom, and p-value. This information can be presented in a table (see for example Table 13.2).

For our example, the results could be presented as follows:

The final multilevel model calculated to test our hypotheses was built up in several steps. Table 13.2 presents the results of the multilevel analyses and model comparisons. The first step involved testing a model without predictors that included a fixed intercept (Model 1). Secondly, a model without predictors including a random intercept was calculated (Model 2) and compared to Model 1 using the chi-square likelihood ratio test. Model 2 showed a bitter fit than Model 1. The analysis was continued by adding self-directedness as a fixed predictor (Model 3). Subsequently, all other predictor variables were added to the model (Model 4). Because self-efficacy ($t = -.52$, df $= 986$, $p = .60$) and pay satisfaction ($t = -.21$, df $= 986$, $p = .84$) were non-significant predictors, the variables were removed from the model (Model 5). After adding all level-1 fixed effects, a random slope model was tested, more specifically a random slope for organisational support was added (Model 6). In a next step the level-2 variable sector was included (Model 7). Finally, a cross-level interaction effect between organisational support and sector was added (Model 8). This did not improve the model fit. Therefore Model 7 was chosen as the final model.

The results of the final model (pseudo-$R^2 = 42.80\%$) show that self-directedness, time management, and employability are significant positive fixed predictors of an employee's learning intention when initial overall differences in learning intention between organisations are taken into account. In addition, results show that employees in the public sector have a higher learning intention than employees in the private sector. Self-directedness is the strongest fixed predictor of all the

Table 13.2 Results multilevel analyses

Model	1	2	3	4	5	6	7	8
Fixed effects – t-statistic (df)								
(Intercept)	107.47***	30.44***	42.81***	44.88***	44.86***	46.58***	38.37***	33.64***
	(992)	(992)	(991)	(986)	(988)	(988)	(988)	(987)
Self-directedness			19.63***	12.96***	13.90***	14.20***	14.42***	14.38***
			(991)	(986)	(988)	(988)	(988)	(987)
Time management				5.26***	5.24***	5.28***	5.31***	5.31***
				(986)	(988)	(988)	(988)	(987)
Employability				3.61***	3.65***	3.70***	3.65***	3.66***
				(986)	(988)	(988)	(988)	(987)
Support				4.14***	4.52***	3.41***	3.50***	2.12*
				(986)	(988)	(988)	(988)	(987)
Self-efficacy				−.52				
				(986)				
Pay satisfaction				−.21				
				(986)				
Sector							−2.67*	−2.02
							(19)	(19)
Support*Sector								.38
								(987)
Random effects – variance components								
(Intercept)		.16	.08	.07	.07	.06	.05	.05
Support						.01	.01	.01
Residual		.77	.57	.52	.52	.52	.52	.52

(continued)

Table 13.2 (continued)

Model	1	2	3	4	5	6	7	8
Model comparison								
−2log likelihood	2,759.26	2,655.01	2,329.42	2,247.70	2,248.04	2,241.80	2,235.96	2,235.81
Models compared		1 vs. 2	2 vs. 3	3 vs. 4	3 vs. 5	5 vs. 6	6 vs. 7	7 vs. 8
χ^2		104.26***	325.59***	81.71***	81.38***	6.24*	5.84*	.15
(df)		(2, 3)	(3, 4)	(4, 9)	(4, 7)	(7, 9)	(9, 10)	(10, 11)

Note: $N_{observations} = 1{,}013$, $N_{groups} = 21$; *$p < .05$, **$p < .01$, ***$p < .001$; Model 1: fixed intercept; Model 2: random intercept; Model 3: adding self-directedness; Model 4: adding subset of fixed predictors; Model 5: Removing non-significant predictors; Model 6: adding random slope; Model 7: adding level-2 variable; Model 8: adding cross-level interaction

hypothesized predictors. Pay satisfaction and self-efficacy were non-significant predictors of an employees' learning attention. Finally, the perceived support of the organisation is a significant positive random predictor of an employee's learning intention. The cross-level interaction between organisational support and sector could however not explain this variation across organisations.

13.4   Applications of HLM in Research on Professional Learning

Throughout this chapter examples from the research field on professional learning have been offered when clarifying the conceptual explanation of HLM. In fact, as already mentioned, all research studies that collected nested data should be analysed with a HLM approach. Within this section, the research topics within the field of professional learning that have been addressed with HLM analysis by prior research will be presented.

When searching for empirical studies within the field of professional learning, a first thing that comes to the fore is that the majority of this type of studies focused on the professional development of teachers who are nested in schools. This can probably be explained by the fact that the related field of research on school effectiveness has a rich tradition of using HLM (e.g., Cools et al. 2008; Opdenakker et al. 2002). Several research studies in the field of teacher development have investigated the effect of professional learning communities (e.g., Chi-Kin Lee et al. 2011; Lakshmanan et al. 2011), the role of teachability culture on collegial trust among teachers (Van Maele and Van Houtte 2011) and the role of teacher learning opportunities for reflective practice (Camburn 2010).

Over the last decade the interest in team learning has grown substantively (see Chap. 36 by Dochy, Gijbels, Raes, and Kyndt). Up till now the majority of the studies that have investigated this topic by aggregating individual scores to the team level (e.g., Raes et al. 2012; Van den Bossche et al. 2006), however recent research demonstrates the relevance of adopting HLM analysis when investigating team learning (Liu and Fu 2011) or advises the use of HLM for future research (Akgün et al. 2007). When considering the adoption of HLM for research on team learning it is important to keep in mind that the outcome variable should be situated at the lowest level of interest.

Other research in the field of professional learning adopting HLM appears to be scarce and scattered. For example, Martin (2009) investigated the role of learning for the motivation and engagement in the workplace using HLM and Yeo and Neal (2004) explored the role of learning orientation for performance in skills acquisition. Xiao (2002) focused on the role of education in salary growth. Other studies have formulated the advice that research on learning climate (Haurer and Westerberg 2012) and transfer of learning (van den Eertwegh et al. 2013) would benefit from adopting HLM.

In sum, research within the field of professional learning adopting HLM, especially research situated outside of the traditional school context is scarce. In addition, it can be noticed that the studies that did adopt HLM are very recent.

13.5 Discussion

Within this final section, the advantages and disadvantages of HLM for research on professional learning will be discussed. The main advantage of HLM is the fact that it is the most appropriate method of analysis for nested data. If the nested structure of the data is ignored, it is more likely that statistical relations are observed in the sample that are in fact not true (Type-1 error), in addition it might be that it is concluded that a relationships holds for individuals when they are actually true for groups (ecological fallacy). HLM allows us to identify and explain the variance at different levels of the data, in other words, it is able to identify the variance at the individual and organisational level. Moreover, predictors at the level of the organisation can be included in the analyses simultaneously with predictors at the level of the individual. Something that can be very important in the research on the professional learning of employees, because an employee's learning is the result of the interplay between the organisation and the individual (Tynjälä 2008).

However, HLM also has disadvantages and limitations. First, fairly large samples are needed, especially when multiple predictors at various levels are included. With small samples, the researcher is often confronted with convergence problems. If the software (e.g., SPSS) presents results after reporting that it was unable to reach convergence, these results should not and in fact cannot be interpreted, therefore other software package (e.g., R) do not report results because it was actually not possible to calculate results. Secondly, HLM is a fairly complex analysis (due to the many possibilities) in comparison with regression analysis, for example it requires the researcher to make decisions throughout the entire process of analysis. Finally, as discussed, it is very difficult to calculate and interpret the effect sizes when conducting HLM, in addition no consensus exists and the existing body of literature rarely reports effect sizes.

This chapter has attempted to introduce the reader with HLM without using mathematical formulas. On the one hand, the aim was to inform the reader on when this method of analysis could be appropriate. On the other hand, the goal was to introduce the reader with how a two-level analysis on cross-sectional data can be executed by integrating a theoretical explanation and illustration of the analysis performed on an authentic dataset and topic within the field of professional learning. It is believed that in a field where employees are nested within organisations and trainees are nested within trainings or trainings institutions, applying HLM can have an added value. With this chapter we hope to make a contribution to the introduction and application of HLM in the field of professional learning.

Appendix: R Code Illustration Learning Intention

```r
##Setting working directory and loading data
setwd("/Users/evakyndt/Publications/Book Chapter HLM")
data<-read.table("Data learning intention.csv",header=TRUE,sep=";")
dim(data)
head(data)


##Installing packages multilevel
install.packages("car");
install.packages("ggplot2");
install.packages("nlme");
install.packages("reshape")
library(car);
library(ggplot2);
library(nlme);
library(reshape)


### data without missing values for multilevel
DataMultilevel1 <-data[, c("OrganisationID","Sector","Intention", "time",
 "efficacy", "directedness", "support", "employability", "financial")]
DataMultilevel <- na.exclude(DataMultilevel1)
dim(DataMultilevel)


##Assessing the need for a multilevel model
interceptOnly <-gls(Intention ~ 1, data=DataMultilevel, method= "ML")
summary(interceptOnly)

randomInterceptOnly <-lme(Intention ~1,data=DataMultilevel, random =
 ~1|OrganisationID, method="ML")
summary(randomInterceptOnly)

anova(interceptOnly, randomInterceptOnly)
logLik(interceptOnly)*-2
logLik(randomInterceptOnly)*-2


##Centring level-1 predictors
DataMultilevel$Ctime <- DataMultilevel$time -mean(DataMultilevel$time)
DataMultilevel$Cfinancial <-DataMultilevel$financial -
 mean(DataMultilevel$financial)
DataMultilevel$Cemployability <- DataMultilevel$employability -
 mean(DataMultilevel$employability)
DataMultilevel$Cefficacy <- DataMultilevel$efficacy -
 mean(DataMultilevel$efficacy)
DataMultilevel$Csupport <- DataMultilevel$support -
 mean(DataMultilevel$support)
DataMultilevel$Cdirectedness <- DataMultilevel$directedness -
 mean(DataMultilevel$directedness)
##Adding fixed effect self-directedness
randomInterceptDirect <-lme(Intention ~ Cdirectedness, data =
 DataMultilevel, random = ~1|OrganisationID, method="ML")
summary(randomInterceptDirect)
anova(randomInterceptOnly, randomInterceptDirect)
logLik(randomInterceptDirect)*-2
```

```
##Adding all predictors
randomInterceptFUll <-lme(Intention ~ Cdirectedness + Ctime + Cefficacy +
 Cemployability + Csupport + Cfinancial, data = DataMultilevel, random =
 ~1|OrganisationID, method="ML")
summary(randomInterceptFUll)
anova(randomInterceptDirect, randomInterceptFUll)
logLik(randomInterceptFUll)*-2


##Removing non-significant predictors
randomInterceptFull2 <-lme(Intention ~ Cdirectedness + Ctime +
 Cemployability + Csupport, data = DataMultilevel, random =
 ~1|OrganisationID, method="ML")
summary(randomInterceptFull2)
anova(randomInterceptFull2, randomInterceptFUll)
logLik(randomInterceptFull2)*-2


##Introducing random slopes for support
addRandomSlope <-lme(Intention ~Cdirectedness + Ctime + Cefficacy +
 Cemployability + Csupport + Cfinancial, data = DataMultilevel, random =
 ~Csupport|OrganisationID, method="ML")
summary(addRandomSlope)
anova(randomInterceptFUll, addRandomSlope)
logLik(addRandomSlope)*-2


##Introducing level-2 predictor
addRandomLevel2 <-lme(Intention ~Cdirectedness + Ctime + Cefficacy +
 Cemployability + Csupport + Cfinancial + Sector, data = DataMultilevel,
 random = ~Csupport|OrganisationID, method="ML", control=list(maxIter=1000,
 msMaxIter=1000, niterEM=1000))
summary(addRandomLevel2)
anova(addRandomSlope, addRandomLevel2)
logLik(addRandomLevel2)*-2


##Introducing cross-level interaction
FinalModel <-lme(Intention ~Cdirectedness + Ctime + Cefficacy +
 Cemployability + Csupport + Cfinancial + Sector + Csupport:Sector, data =
 DataMultilevel, random = ~Csupport|OrganisationID, method="ML",
 control=list(maxIter=1000, msMaxIter=1000, niterEM=1000))
summary(FinalModel)
anova(addRandomLevel2, FinalModel)
```

References

Akgün, A. E., Byrne, J. C., Lynn, G. S., & Keskin, H. (2007). Team stressors, management support, and project and process outcomes in new product development projects. *Technovation, 27*, 628–639.

Anumendem, D. N., Verbeke, G., De Fraine, B., Onghena, P., & Van Damme, J. (2013). Double serial correlation for multilevel growth curve models. *Quality and Quantity, 47*, 1413–1427.

Baert, H., De Rick, K., & Van Valckenborgh, K. (2006). Towards the conceptualisation of learning climate. In R. Vieira de Castro, A. V. Sancho, & P. Guimaraes (Eds.), *Adult education: New routes in a new landscape* (pp. 87–111). Braga: University of Minho.

Billett, S. (2008). Welcome to the journal. *Vocations and Learning, 1*, 1–5.

Camburn, E. M. (2010). Embedded teacher learning opportunities as a site for reflective practice: An exploratory study. *American Journal of Education, 116*, 463–489.

Chi-Kin Lee, J., Zhang, Z., & Yin, H. (2011). A multilevel analysis of the impact of a professional learning community, faculty trust in colleagues and collective efficacy on teacher commitment to students. *Teaching and Teacher Education, 27*, 820–830.

Cools, W., Van den Noortgate, W., & Onghena, P. (2008). ML-DEs: A program for designing efficient multilevel studies. *Behavior Research Methods, 40*, 236–249.

D'Haenens, E., Van Damme, J., & Onghena, P. (2012). Constructing measures for school process variables: The potential of multilevel confirmatory factor analysis. *Quality and Quantity, 46*, 155–188.

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*, 132–150.

De Leeuw, J., & Kreft, I. G. G. (1998). *Introducing multilevel modeling*. London: Sage.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121–138.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading: Addison-Wesley.

Gegenfurtner, A. (2013). Dimensions of motivation to transfer: A longitudinal analysis of their influences on retention, transfer, and attitude change. *Vocations and Learning, 6*, 187–205.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Haurer, E., Nordland, A. M., & Westerberg, K. (2012). Development intention, learning climate and use of knowledge of elderly care. *Journal of Workplace Learning, 24*, 19–33.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370.

Hedges, L. V. (2011). Effect sizes in three-level cluster-randomized experiments. *Journal of Educational and Behavioral Statistics, 36*, 346–380.

Kyndt, E., & Baert, H. (2013). Antecedents of employees' involvement in work-related learning: A systematic review. *Review of Educational Research, 83*, 273–313.

Kyndt, E., Dochy, F., Michielsen, M., & Moeyaert, B. (2009). Employee retention: Organisational and personal perspectives. *Vocations and Learning, 2*, 195–215.

Kyndt, E., Govaerts, N., Dochy, F., & Baert, H. (2011). The learning intention of low-qualified employees: A key for participation in lifelong learning and continuous training. *Vocations and Learning, 4*, 211–229.

Kyndt, E., Dochy, F., Onghena, P., & Baert, H. (2013). The learning intentions of low-qualified employees: A multilevel approach. *Adult Education Quarterly, 63*, 165–189.

Kyndt, E., Onghena, P., Smet, K., & Dochy, F. (Accepted). Employees' learning intention: Comparing low- and high-qualified employees. *International Journal of Educational and Vocational Guidance*.

Lakshmanan, A., Heathe, B. P., Perimutter, A., & Elders, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching, 48*, 534–551.

Liu, D., & Fu, P. (2011). Motivating protégés' personal learning in teams: A multilevel investigation of autonomy support and autonomy orientation. *Journal of Applied Psychology, 96*, 1195–1208.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks: Sage.

Martin, A. J. (2009). Motivation and engagement in the workplace: Examining a multidimensional framework and instrument from a measurement and evaluation perspective. *Measurement and Evaluation in Counselling and Development, 41*, 223–243.

Maurer, T. J., Weiss, E. M., & Barbeite, F. G. (2003). A model of involvement in work-related learning and development activity: The effects of individual, situational, motivational and age variables. *Journal of Applied Psychology, 88*, 707–724.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376–398.

Oosterbeek, H. (1996). A decomposition of training probabilities. *Applied Economics, 29*, 799–805.

Opdenakker, M., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes upon mathematics achievement. *School Effectiveness and School Improvement, 13*, 399–427.

Peugh, J. L. (2010). A practical guide to multilevel modelling. *Journal of School Psychology, 48*, 85–112.

Prinzie, P., Onghena, P., & Hellinckx, W. (2005). Parent and child personality traits and children's externalizing problem behavior from age 4 to 9 years: A cohort-sequential latent growth curve analysis. *Merrill-Palmer Quarterly, 51*, 335–366.

Pustjens, H., Van den Noortgate, W., Onghena, P., & Van Damme, J. (2004). *Multiniveau-analyse in de praktijk. Deel 1: Een eerste introductie. [Multilevel analysis in practice. Part 1: A first introduction]*. Leuven: Acco.

R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Retrieved on July 19, 2012, from http://www.R-project.org/.

Raes, E., Decuyper, S., Lismont, B., Van den Bossche, P., Kyndt, E., Demeyere, S., & Dochy, F. (2012). Facilitating team learning through transformational leadership. *Instructional Science, 41*, 287–305.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Roberts, K. J., & Monaco, J. P. (2006). *Effect size measures for the two-level linear multilevel model*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Simpson, P. A., & Stroh, L. K. (2002). Revisiting gender variation in training. *Feminist Economics, 8*, 21–53.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights: Allyn & Bacon.

Twisk, J. W. R. (2006). *Applied mulitlevel analysis: A practical guide*. Cambridge: Cambridge University Press.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*, 130–154.

Van den Bossche, P., Gijselaers, W., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. Team learning beliefs and behaviour. *Small Group Research, 37*, 490–521.

van den Eertwegh, V., van Duimen, S., Van Dalen, J., Scherpbier, A. J. J. A., & van der Vleuten, C. P. M. (2013). Learning in context: Identifying gaps in research on the transfer of medical communication skills to the clinical practice. *Patient Education and Counselling, 90*, 184–192.

Van den Noortgate, W., & Onghena, P. (2006). Analysing repeated measures data in cognitive research: A comment on regression coefficient analyses. *European Journal of Cognitive Psychology, 18*, 937–952.

Van den Noortgate, W., Opdenakker, M., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement, 16*, 281–303.

Van Maele, D., & Van Houtte, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *American Journal of Education, 117*, 437–464.

Xiao, J. (2002). Determinants of salary growth in Shenzhen, China: An analysis of formal education, on-the-job training, and adult education with a three-level model. *Economics of Education Research, 21*, 557–577.

Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects; of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*, 231–247.