1 Introduction

"Once you know that hierarchies exist, you see them everywhere." I have used this quote by Kreft and de Leeuw (1998) frequently when writing about why, when, and how to use multilevel models and when giving workshops on multilevel modeling. I start this way because I think learning about multilevel modeling meaningfully changes how people think about their research. It can change how they think about data they have already collected (the data are multilevel and this was not recognized or was not taken into account), it can change how they think about the data they will collect, and it can change how they think about the questions they will ask and answer.

It is always frustrating to me when I have a question and perhaps a sense of what type of information (raw data) might be used to answer that question, but I have no idea how I would analyze such data if I collected them. How would I compile the statistics that would answer my question? What statistics would be best? Moreover, I assume that others share this frustration (actually, I know this). For those of you who are not familiar with multilevel modeling, the overarching goal of this volume is to change how you think about your data and the questions they can answer: to expand the tools you have at your disposal so that you can think about different (perhaps better) questions with the knowledge of how you can answer these new questions. Even for those of you who are somewhat familiar with multilevel modeling, some of the techniques I discuss may expand your repertoire.

Certainly, much of this volume will be about the "nuts and bolts" of multilevel modeling. How does one build a multilevel model? How does one test different hypotheses and interpret the results? How on earth do you write this stuff up? And so forth. Nevertheless, the overarching goal is conceptual. The nuts and bolts are just that, simple tools, means to ends. The real prize is the ability to understand phenomena more clearly, to separate relationships that exist at different levels of analysis, and to avoid confounding relationships across levels of analysis.

In service of this conceptual goal, my primary and immediate goal in writing this volume is to help you understand what multilevel modeling is, why it is useful for social and personality psychologists, and how to conduct, interpret, and describe multilevel analyses. The volume is intended for people who know nothing or very little about multilevel analyses and who want to learn how to conduct multilevel analyses, and for people who simply want to read an article that uses multilevel modeling and have some idea of what was done. It is not technically focused; rather, it provides a rationale for multilevel modeling, describes critical features of the technique, and discusses different applications. Throughout, examples are provided that are particularly relevant to social and personality psychologists.

Those of you who are already familiar with multilevel modeling may find that after reading this volume you work more efficiently and with more confidence. In the case of some of the sections that focus on specific aspects of the technique, you may find that they enhance and sharpen your skills. By the way, in terms of statistical background and experience, you will need to be somewhat familiar with regression analysis – simple OLS (ordinary least squares) regression. To take full advantage of some of the "tricks" that I describe, it will be beneficial but not necessary for you to be familiar with different ways of coding grouping variables, dummy codes, contrast codes, etc.

The style of this volume is not the same as that of many contemporary articles in social and personality psychology; each statement I make will not be supported by a slew of references. Points that are not controversial (at least in my mind) will simply be stated as fact. When there is some uncertainty, this will be mentioned. There will be references here and there so that those of you who are starved for citations will find sustenance, however meager it may be. My decision to write this way was motivated by my desire to inform and to instruct while not debating the merits of distinctions that are meaningful only to those who are well informed, however important and necessary such distinctions and debates may be.

Moreover, there will be times when I make recommendations based solely upon my experience: what has happened to my data, what I have found that works or does not work, etc. There are a lot of aspects of MLM that are not well understood, and I am certain that I have missed some recent developments. I am not a statistician *per se*. I am an experienced and (in my own humble opinion) a well-informed user. Regardless, I can assure you that in all the cases I discuss, I have analyzed numerous data sets in different ways, with different options, simply to see what matters and how it matters. I hope that this volume encourages you to do the same. Any statistical procedure is a tool, and understanding what it does can be thought of as a science, but applying it judiciously is an art. So, read this volume and learn how to mix colors so that you will be able to paint your own masterpiece.

Most of the examples I will use will be from my own published research. I do this not because I think my research is better than anyone else's. Rather, I use examples from my own work because I am more familiar with the subtleties of the data and the analyses, and I could use and re-analyze my own data more easily than the data collected by other scholars. Off we go. By the way, I will often use the abbreviation MLM for multilevel modeling.

What is meant by the term "multilevel"?

The term "multilevel" refers to the fact that observations, sometimes called units of analysis, are collected (or sampled) at multiple levels of analysis simultaneously. Okay, that's a bit circular. It may help to note that multilevel data sets are sometimes referred to as "nested" or "hierarchically nested" because observations at one level of analysis are nested within observations at another level. Statisticians also use the phrase "clustered" and sometimes talk about observations being clustered within a common cluster. I will use the term "nested" because its use is more widespread than "clustered."

If for a study of academic achievement the data were collected in 20 classrooms with 10–20 students in each classroom, this would create a multilevel data set, with students nested or clustered within classes. In such a study, data are collected describing units of analysis at different levels of analysis. At the classroom level, class size or years of teacher's experience might be measured. Such measures exist only as classroom characteristics that are shared by all students in a specific classroom. In contrast, at the student level, grades and amount of study, measures that might vary among the students within a single classroom, might also be collected.

Similarly, multiple observations about a single person might be collected in some type of diary study. In such a study, daily observations (e.g., daily observations of mood) would be treated as nested within persons. Person level measures such as gender or some type of personality trait would exist at the person level, and the daily level observations for each person would have these characteristics in common. In contrast, at the daily level, data describing the events that occurred each day and how an individual thought about him or herself that day might be collected. Later, I describe how to conceptualize levels of analysis.

An important aspect of multilevel sampling is that analyses of multilevel data need to take into account the error associated with sampling at multiple levels of analysis. Social and personality psychologists are accustomed to thinking of the sampling error associated with sampling individuals. Nevertheless, in the typical multilevel study a sample of some unit of analysis other than people has been drawn. In a diary study in which the focus of attention is on within-person relationships between daily measures (e.g., relationships between stressful daily events and daily affect), there is some error associated with the sampling of the days over which the study took place. Coefficients describing such relationships that are based on a particular two weeks in a person's life when the study was conducted will typically be similar to, but not the same as, coefficients based on a different two-week period. That is, the coefficients themselves have a sampling error different from, and above and beyond, the sampling error associated with sampling persons. Similarly, when individuals are nested within groups, there are two targets of inference – individuals and groups. The groups in such studies are meant to be representative of the population of groups.

The "multi" in "multilevel" also refers to the fact that relationships and the phenomena they are meant to represent can exist and be examined simultaneously at different levels of analysis (or different levels of aggregation). In the above classroom example, one could examine relationships between amount of study and grades at the between-class level (Are average grades higher in classes in which students study more on average?) or at the within-class or student level (Do students who study more have higher grades?). More subtly, one could determine if individual level relationships between grades and studying varied across classrooms. As discussed below, relationships at these two levels of analysis are mathematically and statistically independent. It is theoretically possible to have positive relationships at one level of analysis and negative relationships at another. Technically, knowing the classroom level relationship between studying and grades tells us nothing about the student level relationship between these same two measures.

Traditionally, the levels of a multilevel model are referred to by number – level 1, level 2, and so forth – with larger numbers indicating levels higher in the hierarchy. So, in the previous example, data describing individual students would be level 1 data (could also be called student level), and data describing classrooms would be level 2 data (could also be called classroom level). In this volume, I will focus on two-level models because they illustrate the principle well and because two levels will be sufficient for most applications. Although the number of levels is theoretically limitless, as discussed below, there are reasons to follow the advice offered by the adage "Less is more."

To me, one of the most powerful advantages obtained by understanding MLM is the fact that the same principles (and techniques) can be applied to data describing vastly different phenomena. Just as we can think of students nested within classrooms, it is only a small step to think of workers nested within work groups or patients/clients nested within therapists, clinics, or treatment centers. Although perhaps not as obvious, as suggested above, it is just another small step to think of diary data in which observations (e.g., daily reports) are nested within persons.

The same modeling techniques hold whether people are nested within groups or observations are nested within persons. Admittedly, there are some concerns that are more important for one broad type of data than for others. For example, autocorrelated errors (the possibility that errors of measurement for observations collected over time are correlated) might be a concern when analyzing some types of diary data, whereas they would not be a concern for a study in which people were nested within groups. Nevertheless, there are more similarities among the MLM procedures appropriate for substantively different data structures than there are differences among such procedures. In learning terms, there is considerable positive transfer of understanding the techniques needed to analyze the data in one substantive domain to another.

NB: most multilevel modelers use the term “group” to refer to an organizing unit, even when an organizing unit is not an actual group. For example, in a diary study in which days are nested within persons, within the lexicon of MLM, people are referred to as groups. Admittedly, referring to individuals as groups can be confusing, but within MLM the use of the term “group” in this way is so deeply ingrained that it is unavoidable at present. Perhaps with time the term “cluster” will replace “group,” but for now we will have to grin and bear it.

Varying relationships across different levels of analysis

The importance of taking into account the nested or multilevel structure of a data set reflects (in part) the possibility, perhaps the likelihood, that relationships between constructs at different levels of analysis vary. For example, assume we have collected data describing how much students, who are nested within classrooms, study, and we also know their grades. One way to analyze such data would be to calculate for each classroom the average amount students study each week and the average grade they receive, and correlate these two measures. This is sometimes referred to as an analysis of aggregates. Such a correlation would answer the question, “Are grades higher in classes in which students study more compared to classes in which students study less?”

This is an appropriate question, but it is not the same question as, “Do students who study more get higher grades?” This second question refers to a relationship at the level of the individual student, not at the level of the classroom. As illustrated by the data in Table 1, it is entirely possible to have one type of relationship (positive v. negative) at one level of analysis and another relationship at the other level of analysis. In the first panel, the relationship between grades and studying within each class is negative, whereas the relationship at the between-class level (between class averages) is positive. In contrast, in the second panel, the relationship between grades and studying within each class is positive, whereas the relationship at the between-class level is negative. In the third panel, the within-class relationships vary. Relationships at the two levels of analysis are mathematically independent. Knowing the relationship at one level of analysis tells us nothing (technically speaking) about relationships at the other level of analysis.

In case you are having trouble thinking about how a correlation between studying and grades at the between-class level does not accurately represent relationships at the student level, consider the following example. In the clinical literature, anxiety and depression are “comorbid.” People who are anxious tend to be depressed and vice-versa. Moreover, such a relationship is assumed by most

Table 1 Varying relationships across levels of analyses

Panel 1 Negative within-class relationship, positive between-class relationship						
	Class 1		*Class 2*		*Class 3*	
	Grades	*Study*	*Grades*	*Study*	*Grades*	*Study*
	2.1	6	3.1	8	3.3	10
	2.2	5	3.2	7	3.4	9
	2.3	4	3.3	6	3.5	8
	2.4	3	3.4	5	3.6	7
	2.5	2	3.5	4	3.7	6
Mean	2.3	4	3.3	6	3.5	8
Panel 2 Positive within-class relationship, negative between-class relationship						
	Class 1		*Class 2*		*Class 3*	
	Grades	*Study*	*Grades*	*Study*	*Grades*	*Study*
	2.1	6	3.1	4	3.3	2
	2.2	7	3.2	5	3.4	3
	2.3	8	3.3	6	3.5	4
	2.4	9	3.4	7	3.6	5
	2.5	10	3.5	8	3.7	6
Mean	2.3	8	3.3	6	3.5	4
Panel 3 Variable within-class relationship, positive between-class relationship						
	Class 1		*Class 2*		*Class 3*	
	Grades	*Study*	*Grades*	*Study*	*Grades*	*Study*
	2.1	2	3.1	8	3.3	10
	2.2	3	3.2	5	3.4	9
	2.3	4	3.3	6	3.5	8
	2.4	5	3.4	7	3.6	7
	2.5	6	3.5	4	3.7	6
Mean	2.3	4	3.3	6	3.5	8

measures of the Big Five factor of neuroticism. Most measures of neuroticism have items such as “depressed, blue” and “gets nervous easily.” Nevertheless, at any moment in time, it may be difficult for people to be both depressed and anxious because depression is a type of deactive affect, whereas anxiety is a type of active affect. At the within-person (moment to moment) level, depression and anxiety may be negatively related or unrelated, whereas at the between-person level (how depressed and anxious a person is, in general, on average), the two may be positively related. See Cervone (2004) and Affleck, Zautra, Tennen, and Armeli (1999) for discussions of the value of distinguishing between- and within-person levels of analysis when considering individual differences.

Returning to our classroom example, as indicated by the data in Table 1, within-group relationships (for our example, the individual or within-classroom level) can vary, and one of the advantages of the techniques discussed in this volume is the ability to model such variability. Why is the relationship between grades and studying stronger in some classes than in others? Similarly, at the within-person level, relationships between measures can vary. Why is the within-person relationship between daily stressors and anxiety stronger for some people than for others? One of the important advantages of multilevel analyses over single level analyses is that they allow for the possibility that relationships between measures vary across units of analysis (groups at level 2). Moreover, MLM provides statistically accurate and efficient estimates of how between-unit (level 2) differences can account for within-unit (level 1) differences in relationships.

When relationships vary across levels of analysis or across units within the same level of analysis, such situations beg questions about which is the "correct" relationship. What's the right answer? Unfortunately, there is no simple answer to such questions. The correct answer depends upon the question. If the question concerns between-group relationships, then analyses at the between-group level provide the answer. If the question concerns within-group relationships, then analyses that describe within-group relationships provide the answer, with the caveat that within-group relationships may vary.

Different ways of analyzing multilevel data

Over time, nested data have been analyzed in various ways other than using the techniques I discuss here (what are technically referred to as multilevel random coefficient models – see next chapter), and in this section I critically review these approaches. Most of these previous methods rely on some type of OLS analysis, and although with the increasing popularity of multilevel modeling these types of analyses are appearing less and less often, as Santayana warned, "Those who cannot remember the past are condemned to repeat it," and so it will be instructive to review briefly such approaches.

Broadly speaking, in the past, multilevel data have sometimes been analyzed with what have been called "aggregation" and "disaggregation" techniques. In aggregation analyses, group means are calculated and relationships are examined at the between-group level. Although aggregation analyses can be appropriate (depending upon the level at which a question is posed), researchers who rely on them are prone to commit what is commonly referred to as the ecological fallacy (Robinson, 1950). Researchers commit the ecological fallacy when they use relationships at the between-group level to draw conclusions about relationships at the within-group level. Robinson's classic paper was based on analyses of the 1930 US Census. Using aggregates calculated within each state, he found a positive between-state relationship between literacy rates and the percentage of residents that were immigrants. States that had more immigrants had higher literacy rates. In contrast, he found negative relationships within states, i.e., the literacy rate among immigrants was lower than it was among those who were native born. Such possibilities are also illustrated by the data presented in Table 1, particularly the data in the first two panels.

In disaggregation analyses, analyses are conducted at level 1 (in a two-level data set). In such analyses, level 2 measures are "brought down" to level 1 (level 2 measures are repeated for all the level 1 units nested within their corresponding level 2 unit) and treated as if they were level 1 measures. In a diary study, this would entail assigning individual differences such as personality measures with each day of data a person provided. In a group study, it would entail assigning group level measures to all of the individuals in a group.

An important characteristic of nested data is that level 1 observations are not fully independent. The members of a group have the characteristics of their group in common, and the social interactions a person describes have the characteristics of the person describing them in common. Such a lack of independence means that techniques such as OLS regression in which level 1 observations are the sole units of analysis cannot be used because such analyses violate a fundamental assumption of such analyses – the independence of observations. In a study of groups, it is incorrect to append group level data to each of the individuals in a group and then conduct a single level analysis with the individuals as the unit of analysis. Likewise, in a diary study, it is incorrect to append individual (person level) data such as personality characteristics to the daily diary data and then conduct single level analyses with the day as the unit of analysis.

Such analyses have other important shortcomings. Of particular importance is that they assume that the level 1 (within-unit) relationships are consistent across level 2 units. This is even the case if a least-squared dummy variable analysis is used (LSDV: e.g., Cohen & Cohen, 1983). In LSDV analyses, the group membership of the level 1 units is represented by a series of $k - 1$ dummy-coded (0, 1) variables, where k is the number of groups. LSDV analyses do take into account the possibility (actually, the likelihood) that the means of level 1 predictors vary across level 2 units, but they do not take into account the possibility (again, the likelihood) that relationships between these level 1 variables vary across level 2 units. Such a possibility can be addressed by including terms representing the interaction between the group variables (the dummy codes) and the various predictors. Nonetheless, aside from practical considerations (e.g., 100 level 2 units and two predictors would require just under 200 level 1 predictors), as discussed above, such analyses do not model error properly. In a multilevel study there are two sampling distributions, and because LSDV analyses are OLS, they can have only one error term. See Nezlek (2001) for a discussion of the shortcomings of various types of OLS analyses of multilevel data.