Human Communication Research ISSN 0360-3989

ORIGINAL ARTICLE

A Primer on Multilevel Modeling

Andrew F. Hayes

School of Communication, The Ohio State University, Columbus, OH 43210

Multilevel modeling (MLM) is growing in use throughout the social sciences. Although daunting from a mathematical perspective, MLM is relatively easy to employ once some basic concepts are understood. In this article, I present a primer on MLM, describing some of these principles and applying them to the analysis of a multilevel data set on doctor–patient communication during medical consultations.

doi:10.1111/j.1468-2958.2006.00281.x

Over the years, there have been at least two orchestrated calls in the field of communication for our thinking about process and theory to acknowledge more than it does that communication is ultimately a multilevel phenomenon. These calls have taken the form of two special issues related to the topic, one in the *Journal of Communication* in 1983 and the other in *Communication Research* in 1991. Contributors to these special issues have lamented researchers' tendency to focus on one level of analysis, thereby producing, "theoretical islands" isolated from one another by "seas of evidence" (Pan & McLeod, 1991, p. 141).

Whether these special issues had any impact on the field is a matter of debate. Let us assume that some progress has been made, but that for the most part the field continues to exist as islands of researchers with their theories, aware of but lacking interest in the residents of neighboring islands. At that time, Ritchie and Price (1991) speculated that the failure of communication researchers to cross levels of analysis—their apparent inability to or disinterest in parting the seas—might be attributable at least in part to the fact that "readily available … statistical techniques are not conducive to cross-level research" (p. 138). This explanation seems wanting, both then and now. Pan and McLeod (1991) pointed out that statistical methods already existed that allow researchers to examine the independent and interactive effects of variables measured at different levels on communication-related outcomes—contextual analysis and multilevel modeling (MLM), for instance. Perhaps, then, was it the lack of implementation of these methods in readily available software? Perhaps, but today, these statistical tools have become even more developed and available to the user through their implementation in software that can be used right on the

Corresponding author: Andrew F. Hayes; e-mail: hayes.338@osu.edu

researcher's desktop computer. So why do we still see relatively little multilevel research in the field's journals?

To me, a plausible explanation is a lack of *awareness* of the statistical tools available rather than their lack of availability, combined with a dearth of good examples of multilevel research and analysis. Although some may argue that the theoretical horse should pull the statistical cart, I argue that the horse and cart are mutually interdependent and should not be separated or ordered in terms of importance to the research enterprise. Just as statistical techniques can help us to answer our theoretical questions, they can also contribute to the formulation of the substantive and theoretical questions we ponder. In other words, knowing what is possible analytically can influence how we think theoretically. It is this belief that motivated this third call for a special issue on multilevel thinking, this time with a focus on multilevel *analysis* rather than just multilevel thought. By increasing awareness of what can be done statistically while simultaneously providing models of people actually using those tools, communication researchers will be prompted to think differently next time they design a study and hopefully will be inspired to try their own multilevel analysis.

In this special issue, you will find several examples of MLM in action. Although the authors were asked to provide more detail than would be typical for a research article published in a regular issue of *Human Communication Research*, they obviously could not go so far as to explain the basic principles and logic of MLM. Instead, that is my job and that is what I attempt to do here in the form of a brief primer. Although MLM is a computationally intense method, the mathematics occurs behind the scenes and need not concern the user who has a good grasp of some basic concepts and can apply them correctly when setting up and interpreting a multilevel model. So in this primer, I introduce some of the important concepts in MLM and show how substantive questions can be answered with the method by working through a set of examples, focusing on only some of the more simple types of multilevel analyses that are possible. This should be thought of as a broad conceptual overview rather than a comprehensive guide, to be used as a springboard for jumping into some of the literature on MLM. Throughout, I cite specific pages of many of the good books available as places to go to for more in-depth discussions of some of the complexities, controversies, and issues I just gloss over.

Although there are many programs that can estimate multilevel models, I use SPSS in this primer because it is available to most communication researchers at little or no cost through institutional site licenses. As most readers will already be familiar with the user interface, focusing on SPSS eliminates this potential distraction to the process of learning how to conduct a MLM. To be sure, SPSS does have a number of limitations, such as its exclusive implementation of the Wald test for testing variance parameters and its inability (as of the date of this writing) to model binary or discrete, ordinal outcomes. Other programs capable of conducting these analyses include SAS, HLM, Mplus, S-plus, and MLwiN, to name just a few. HLM is perhaps the easiest of the specialized programs to use, but it is designed to do nothing other

than MLM. Mplus is one good program that offers a breadth of state-of-the-art methods of analyses such as structural equation modeling in addition to MLM while still having tremendous power for conducting complex multilevel models. Specialized programs often have better computational algorithms and are more likely to produce a solution than general data analysis programs such as SPSS. If you cannot get a model to estimate in SPSS, it may be estimated in a specialized multilevel analysis program. Indeed, once you are comfortable with MLM, you may find yourself moving to one of these more specialized programs.

Multilevel modeling concepts and implementation

The data used throughout this primer are inspired by Cegala, Street, and Clinch (in press) on the effects of patient participation during doctor–patient consultation on information transmission during the medical interview.[1] Four hundred and fifty doctor–patient interactions were observed, with the 450 patients spread across 48 doctors. Thus, patients are *nested* within doctors, with each patient seeing only a single doctor but each doctor seeing anywhere between 4 and 14 patients in the study. Such nesting is a standard feature of multilevel data. Each patient nested under the same doctor is likely to be influenced similarly by processes and attributes that are characteristics of the doctor (e.g., the doctor's approach to medicine, how large his or her practice is, how well the doctor communicates, whether the doctor practices alone or with other doctors, etc.). This feature of multilevel data vastly widens the scope and nature of questions that can be answered. At the same time, such nesting can invalidate the use of "single-level" ordinary least squares regression, analysis of variance, and other forms of analysis that fail to account for the nonindependence between observations that such nesting produces (Bliese & Hanges, 2004; Kenny & Judd, 1986).

Each of the consultations was recorded and transcribed, and each utterance (essentially a "thought unit") was coded into one of several categories. The primary independent variable of interest is *patient participation*, operationalized as the sum of the number of utterances from the patient coded as information seeking, assertion, providing information, or expression of concern. The outcome variable is how much information the doctor *volunteered*, quantified as the number of the doctor's utterances that were not in response to direct questions or comments that the patient made during the consultation. Another variable used as a statistical control in some of the analyses described below is the length of the consultation, in minutes. These are all *level-1* variables, as they are measured attributes of specific doctor–patient interactions. Finally, *doctor busyness* was measured at a higher level, level-2, operationalized as the average number of patients the doctor sees on a daily basis. This is a level-2 variable because it is an attribute of the level-2 unit (doctor) under which the level-1 units are nested. By definition, a level-2 variable does not vary between the level-1 units nested under the same level-2 unit. Level-1 variables, in contrast, typically vary between level-1 units nested under the same level-2 unit.

A screenshot of part of the SPSS data file can be found in Figure 1, illustrating how a multilevel data set should be configured in SPSS for MLM. Each row of the data file contains the data from a specific doctor–patient interaction, and each column contains measurements on each of the variables. As can be seen, this data file looks like one in SPSS would typically look. The distinguishing feature here is an additional variable that codes which level-2 unit (doctor) each level-1 unit (patient) is nested under. The variable name is "docid" for "doctor identification." So the first 12 rows correspond to the 12 patients who consulted with doctor 1, the next 12 rows correspond to patients who consulted with doctor 2, and so forth. The values under "volpro" and "ppart" are the number of utterances volunteered by the doctor and the amount the patient participated during the consultation represented in that row, respectively. "Length" is the length of that consultation in minutes. The column labeled "busy" is a level-2 variable, the average number of patients a doctor sees in

doctor_patient.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 :

	docid	patient	volpro	ppart	length	busy	ppartc	lengthc	busyc
1	1	1.00	134.00	29.00	35.00	16.16	-3.33	14.28	-5.89
2	1	2.00	107.00	32.00	16.00	16.16	-.33	-4.72	-5.89
3	1	3.00	55.00	27.00	13.00	16.16	-5.33	-7.72	-5.89
4	1	4.00	122.00	34.00	36.00	16.16	1.67	15.28	-5.89
5	1	5.00	135.00	33.00	19.00	16.16	.67	-1.72	-5.89
6	1	6.00	178.00	42.00	24.00	16.16	9.67	3.28	-5.89
7	1	7.00	134.00	27.00	33.00	16.16	-5.33	12.28	-5.89
8	1	8.00	107.00	30.00	17.00	16.16	-2.33	-3.72	-5.89
9	1	9.00	55.00	26.00	14.00	16.16	-6.33	-6.72	-5.89
10	1	10.00	122.00	36.00	35.00	16.16	3.67	14.28	-5.89
11	1	11.00	135.00	33.00	19.00	16.16	.67	-1.72	-5.89
12	1	12.00	178.00	41.00	27.00	16.16	8.67	6.28	-5.89
13	2	13.00	107.00	56.00	25.00	24.08	23.67	4.28	2.03
14	2	14.00	80.00	37.00	33.00	24.08	4.67	12.28	2.03
15	2	15.00	15.00	23.00	9.00	24.08	-9.33	-11.72	2.03
16	2	16.00	53.00	23.00	41.00	24.08	-9.33	20.28	2.03
17	2	17.00	49.00	48.00	19.00	24.08	15.67	-1.72	2.03
18	2	18.00	47.00	28.00	13.00	24.08	-4.33	-7.72	2.03
19	2	19.00	107.00	57.00	27.00	24.08	24.67	6.28	2.03
20	2	20.00	80.00	36.00	33.00	24.08	3.67	12.28	2.03
21	2	21.00	15.00	23.00	10.00	24.08	-9.33	-10.72	2.03
22	2	22.00	53.00	23.00	43.00	24.08	-9.33	22.28	2.03
23	2	23.00	49.00	50.00	22.00	24.08	17.67	1.28	2.03
24	2	24.00	47.00	27.00	12.00	24.08	-5.33	-8.72	2.03
25	3	25.00	9.00	39.00	13.00	20.03	6.67	-7.72	-2.03
26	3	26.00	71.00	24.00	12.00	20.03	-8.33	-8.72	-2.03

Figure 1 An SPSS data file screenshot illustrating a multilevel data set.

a day. Notice it is repeated for all rows corresponding to that doctor because it is a characteristic of the doctor—how busy he or she tends to be—and does not vary between patients who consulted the same doctor during the period of study. In SPSS, no special designation is necessary to distinguish between level-1 and level-2 variables in the data file—they all get their own column.[2]

At its core, MLM is not that much different from ordinary "single-level" regression with which most communication researchers are familiar. In MLM, an outcome variable is modeled as a linear combination of predictor variables, each weighted by a coefficient that quantifies how variation in that predictor variable is related to variation in the outcome. One of the primary differences between MLM and other forms of analysis is the ability to estimate one or more of the coefficients or "effects" in the model as either *fixed* or *random*. A fixed effect has only a single value in the model and is applied to each level-1 unit in the analysis regardless of the level-2 unit under which a case is nested. A random effect, in contrast, is allowed to vary between level-2 units. In ordinary regression, the regression intercept and the regression coefficients are both fixed effects, and the regression residual is a random effect. In MLM, we can have several regression intercepts for the model and several regression coefficients for the same predictor, one for each level-2 unit, by setting some or all of the coefficients as random. This is part of what gives MLM its versatility.

The choice as to whether to set an effect as fixed or random is not always an easy one to make. If it makes sense to assume or predict on theoretical or methodological grounds that the relationship between a level-1 variable and the outcome differs between level-2 units, this suggests setting the effect as random. Such an assumption can always be tested, and if the data are inconsistent with this assumption, the model can be reestimated setting the effect to fixed. Almost always, the intercept in a multilevel model is estimated as a random effect, for this allows the mean of the level-1 units nested under a common level-2 unit to vary between level-2 units. It is such variation between level-2 units that accounts for much (but not necessarily all) of the nonindependence between level-1 units in data that are nested, nonindependence that can invalidate the use of single-level methods of analysis such as ordinary multiple regression. For guidance on how to make the choice between estimating an effect as fixed or random, see Snijders and Bosker (1999, pp. 41–45, 80) and Jackson and Brashers (1994, pp. 4–8).

The distinction between a random and a fixed effect is perhaps one of the most difficult ideas to master in MLM and is likely to cause some confusion at first, in part because of our tendency to equate the term "random" with concepts such as "unsystematic," "chance," or "unexplained." A random effect is not necessarily an unsystematic one. It may be that the variation we see in a random effect across level-2 units is very systematic. Indeed, it may be our goal to seek out and then explain the variation in a random effect in a multilevel model. Although this is not standard terminology, I believe it is easier when conceptualizing and interpreting a multilevel model to think in terms of *fixed components* and *random components*. A fixed effect has only a fixed component, a value for that effect that is applied to all level-1 units in

the analysis regardless of which level-2 unit the case is nested under. A random effect usually has both a fixed component and a random component. The fixed component applies to all level-1 units, whereas the random component "customizes" the effect to that level-2 unit by expressing it as a deviation between the effect in that level-2 unit and the effect represented by the fixed component. We often are interested not in the values of those deviations per se but whether and by how much those random components vary between level-2 units. So when we interpret a random effect, we usually focus partly on the fixed component and partly on the variance of the random component.

MLM is a "large sample" procedure, meaning that the mathematics and the theoretical assumptions that underlie the statistics produced by a multilevel program are based on the behavior of statistics calculated in large samples. At issue is the validity of the inferential tests as well as the power of those tests to detect an effect deemed worthy of detecting. As always, just how large is large enough is a question difficult to answer, as it depends on so many things. Further complicating the matter, required sample size is a function of both the number of level-1 units and the number of level-2 units. Ideally, the investigator will have many level-1 units nested under many level-2 units. All other things being equal, power is highest when both level-1 and level-2 sample sizes are large, and for this reason, I recommend that you devote your resources accordingly to maximize both in so far as possible. In addition to the power advantages large samples afford, some programs, SPSS included, have trouble in estimating models when the sample size is too small. The computation of power and required sample size in MLM is a burgeoning area, mathematically complicated and, as always, often requires more information than the investigator likely has available to him or her before the data are collected. Hox (2002, pp. 173–196), Snijders and Bosker (1999, pp. 140–152), and Kreft and de Leeuw (1998, pp. 119–126) provide good treatments on power in MLM. Moerbeek, van Breukelen, and Berger (2000), Raudenbush (1997), and Jackson and Brashers (1994) discuss design issues, including sample size and power, in the planning of multilevel experiments.

Fitting multilevel models

Having introduced some of the important concepts, I next turn to using MLM to answer questions about the relationship between patient participation and how much information a doctor volunteers to a patient during a medical interview. I start by first covering the most basic model that ascertains only the extent to which doctors differ in how much they volunteer. I then estimate models of increasing complexity to answer different questions, ending with a relatively complicated model that estimates how much doctors volunteer to their patients as a function of patient participation, length of the interview, how busy the doctor is on a typical day, and how time available for the consultation interacts with patient behavior during the interview. Although with today's software MLM is literally a point-and-click away, as you will see, setting up a multilevel model does require careful thought in order to assure that the hypothesis of interest is actually being tested with the model specified.

Do doctors differ, on average, in how much information they volunteer?
The first model typically fit when estimating a multilevel model is called the "baseline," "null," or "empty" model. This is a model without any predictors whatsoever, and it focuses on assessing whether the level-2 units (doctors in this case) differ from each other, on average, on the outcome variable. It is equivalent to a one-way analysis of variance, but the group factor is treated as a random effect. In *multilevel equations form*, the model is

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + r_{ij}$$
$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}$$

where Y_{ij} is how much information doctor j volunteers to patient i during the medical interview. The analyses I describe here treat Y as a continuous variable measured at the interval level or higher. Much of the mathematical theory of MLM has been based on this assumption, but it is possible to model outcome variables that are ordinal or dichotomous (see e.g., Guo & Zhao, 2000; Hox, 2002, pp. 103–122; Raudenbush & Bryk, 2002, pp. 291–335; Snijders & Bosker, 1999, pp. 207–238). In the level-1 model above, β_{0j} is the average amount of information doctor j volunteers across all his or her patients, and r_{ij} quantifies the difference between doctor j's average and how much this doctor volunteers to patient i. In the level-2 model, γ_{00} is the grand mean—the average amount doctors volunteer to their patients (aggregating across doctors). Finally, u_{0j} is the difference between doctor j's average and this grand mean.

Written in *mixed model* form by substitution of the level-2 equation into the level-1 equation, the model is

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

This form of the model shows that how much doctor j volunteers to patient i is a function of three components: how much doctors volunteer on average (γ_{00}), how much doctor j's mean differs from this grand mean (u_{0j}), and how much doctor j volunteers to patient i during the consultation differs from his or her own average (r_{ij}). This is a *random intercept–only model*, with the quantity $\gamma_{00} + u_{0j}$ being the random intercept containing a fixed component (γ_{00}) and a random component (u_{0j}). The last component, r_{ij}, is the level-1 residual, also a random effect. Notice that the random components in the model contain a j subscript, whereas the fixed component does not, a pattern that is consistent across all models discussed in this paper. The algorithms and hypothesis tests assume that the random component of the intercept is distributed normally with mean zero and variance τ_{00} and that the residuals are normally distributed with mean zero and variance σ^2. Some of the interpretation of a multilevel model focuses on the estimates of these variances.

Figure 2 presents the SPSS output from this model, and a summary of this output and all other models discussed in this paper can be found in Table 1. In Appendix A, the SPSS syntax used to estimate all the models described in this paper

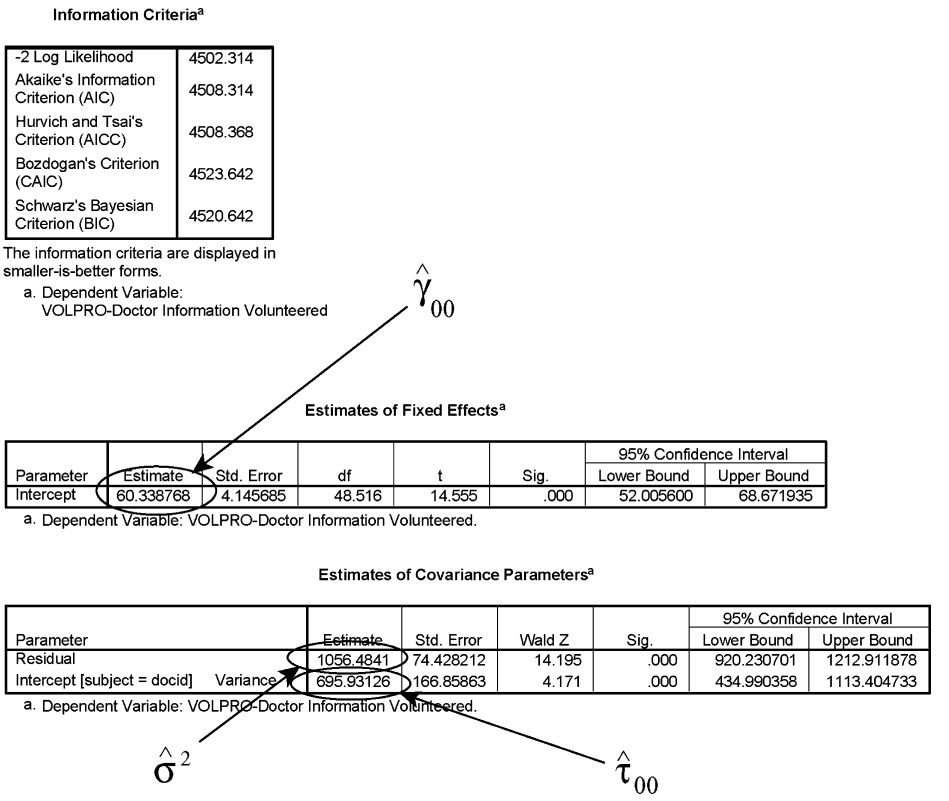


Figure 2 SPSS MIXED output for Model 1.

is provided. In all models, maximum likelihood (ML) estimation was used. Restricted maximum likelihood estimation (REML) is also available in SPSS and other programs. The choice between ML and REML tends not to matter much unless the number of level-2 units is small or when the number of fixed effects is large, in which case REML is preferred. For a discussion, see Snijders and Bosker (1999, pp. 56–57) and Kreft and de Leeuw (1998, pp. 130–135).

From Figure 2 and Table 1 (Model 1), $\hat{\gamma}_{00} = 60.339$, meaning that, on average, doctors volunteered 60.339 utterances. Although the test of significance shows this average is statistically different from zero, $t(49) = 14.555$, $p < .001$, this is not particularly interesting or informative.[3] More important is whether there is evidence that the doctors differed from each other in how much information they provided, on average. This question is tested by assessing whether the variance of the random components of the intercept—the u_{0j} in the earlier equations—is different from zero. If doctors differed very little from each other, then the j values of u_0 should differ little from each other and thus exhibit little to no variance. The estimated variance of the random components (i.e., the u_{0j}s), represented symbolically as $\hat{\tau}_{00}$, is 695.931

Table 1 Parameter Estimates for the Seven Models Examining the Relationship Between Patient Participation and How Much Information a Doctor Volunteers During a Consultation

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Fixed components								
Intercept	$\hat{\gamma}_{00}$	60.339***	13.179***	59.403***	58.567***	59.564***	74.254***	59.581***
PPART	$\hat{\gamma}_{10}$		1.430***					
PPARTC	$\hat{\gamma}_{10}$			1.430***	0.968***	0.989**	3.689*	0.979**
LENGTHC	$\hat{\gamma}_{20}$				0.800***	0.864***	0.865***	0.865***
BUSY	$\hat{\gamma}_{01}$						−0.668	
BUSYC	$\hat{\gamma}_{01}$							−0.668
BUSY × PPARTC	$\hat{\gamma}_{11}$						−0.123*	
BUSY × PPARTC	$\hat{\gamma}_{11}$							−0.123*
Variance of random components								
	$\hat{\tau}_{00}$	695.931***	705.748***	705.748***	762.921***	797.469***	810.694***	810.694***
	$\hat{\tau}_{11}$					1.954*	1.704*	1.704*
	$\hat{\tau}_{01}$					17.790+	16.172†	16.172+
	$\hat{\sigma}^2$	1,056.484	921.134	921.134	840.139	730.020	722.065	722.065
Deviance (−2LL)		4,502.314	4,446.808	4,446.808	4,412.499	4,389.625	4,383.537	4,383.537

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

and statistically different from zero, $Z = 4.171$, $p < .001$. Taking the square root of this variance, we can interpret this as the estimated standard deviation of the doctor means.

This Wald test is only one way of testing the significance of a random effect. As Hox (2002, pp. 148–149) points out, this test assumes the sampling distribution of the variance is normal, an assumption that can be unreasonable (Raudenbush & Bryk, 2002, p. 64). An alternative is a *likelihood ratio test*, conducting by comparing the "deviances" of two models, one in which the effect of interest is fixed and other in which it is allowed to vary randomly across level-2 units. In the SPSS syntax above, the doctor-level intercept is set to fixed by eliminating the "RANDOM" subcommand. In SPSS, the deviances are printed in the output as the "−2 Log Likelihood" (−2LL). In the model with the random intercept, −2LL = 4,502.314. Without the random component of the intercept, −2LL = 4,643.326, a difference of 4,643.326 − 4,502.314 = 141.012. These two models differ by one parameter estimate, the intercept variance $\hat{\tau}_{00}$. As a result, the difference in the deviances is distributed as χ^2 with a single degree of freedom under the null hypothesis that the variance of the random intercept components is zero. For $\chi^2(1) = 141.012$, the p value is less than .001, leading a rejection of this null hypothesis.[4] In general, the likelihood ratio test is a better test, and it can conflict with the Wald test. In such cases (an example of which is presented later), trust the likelihood ratio test more. Unfortunately, SPSS will not conduct the likelihood ratio test automatically, meaning you will have to conduct the test manually as I have. SPSS only prints the results of the Wald test.

The null model is typically estimated first not only to assess whether the level-2 units differ on the outcome but also in order to estimate the degree of nonindependence in the outcome variable across level-1 units. This estimate is called the *intraclass correlation* or ICC and is defined as

$$\text{ICC} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2}$$

where $\hat{\sigma}^2$ is the estimated residual variance (the variance of the values of r_{ij}), found in the SPSS output in the "Residual" row. From Figure 2, $\hat{\sigma}^2 = 1,056.484$ and $\hat{\tau}_{00} = 695.931$, so ICC = 695.931/(695.931 + 1,056.484) = 0.397. We can say that 39.7% of the total variance in the number of utterances doctors volunteered when interacting with their patients is accounted for by differences between doctors in how much they volunteer on average. Some argue that MLM is not necessary if ICC is sufficiently close to zero, as this implies that level-1 units (patients) are statistically independent. However, just how close to zero is "sufficiently close" depends on a number of things, and values of ICC as small as 0.05 can invalidate hypotheses tests and confidence intervals when MLM is not used (see, e.g., Kreft & de Leeuw, 1998, pp. 9–10). Furthermore, there are benefits to the use of MLM even when the ICC is near zero.

The residual variance from the null model is also important information because it gives us a baseline for estimating the statistical importance of level-1 predictors in a multilevel model, something illustrated later.

Examining the relationship between patient participation and how much information a doctor volunteers

If patient participation prompts doctors to volunteer more information, we would expect that a doctor would volunteer more information to a patient who participates relatively more. To determine if so, we estimate how much information doctor j volunteers to patient i from information about how much patient i participates. We still allow doctors to vary in how much they volunteer on average by estimating the intercept in the model as a random effect. The multilevel equations for this model are

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}\text{PPART}_{ij} + r_{ij}$$
$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j};\ \beta_{1j} = \gamma_{10}$$

where β_{1j} quantifies the relationship between how much information doctor j volunteers to his or her patients as a function of how much those patients volunteer. β_{1j} is conceptually equivalent to the unstandardized regression coefficient in single-level regression.[5] The mixed model form produced by substituting the level-2 model into the level-1 model is

$$Y_{ij} = \gamma_{00} + \gamma_{10}\text{PPART}_{ij} + u_{0j} + r_{ij}$$

Notice that there are no us in the level-2 model of β_{1j}. As a result, the effect of participation is fixed as constant across doctors, with that effect quantified as γ_{10}. That is, all j values of β_1 are fixed to be equal to γ_{10}, meaning that there is really only a single coefficient estimating the effect of patient participation. Thus, patient participation is a fixed effect with only a fixed component, whereas the intercept is a random effect with both a fixed and a random component (the u_{0j} component in the level-2 model of β_{0j}). Fixing β_{1j} to a constant value γ_{10} is an assumption being made now that is tested later. Of course, by allowing the intercept to vary randomly, no assumption is being made that all of the variation in how much information doctors volunteer is accounted for by patient participation. We can also test this assumption from information the model yields.

Figure 3 and Table 1 (Model 2) provide the relevant output from SPSS. Our attention immediately focuses on $\hat{\gamma}_{10}$, the estimate for the effect of patient participation. That estimate is displayed as $\hat{\gamma}_{10} = \hat{\beta}_{1j} = 1.430$, $t(444) = 7.712$, $p < .001$, leading to a rejection of the null hypothesis of no relationship between patient participation and how much information a doctor volunteers. The interpretation is that two patients who differ by one unit in their participation are estimated to differ by 1.430 in the number of utterances their doctors volunteer, with the patient participating more receiving more volunteered information from the doctor. This test can also be conducted by comparing the deviance of the null model (Model 1) to

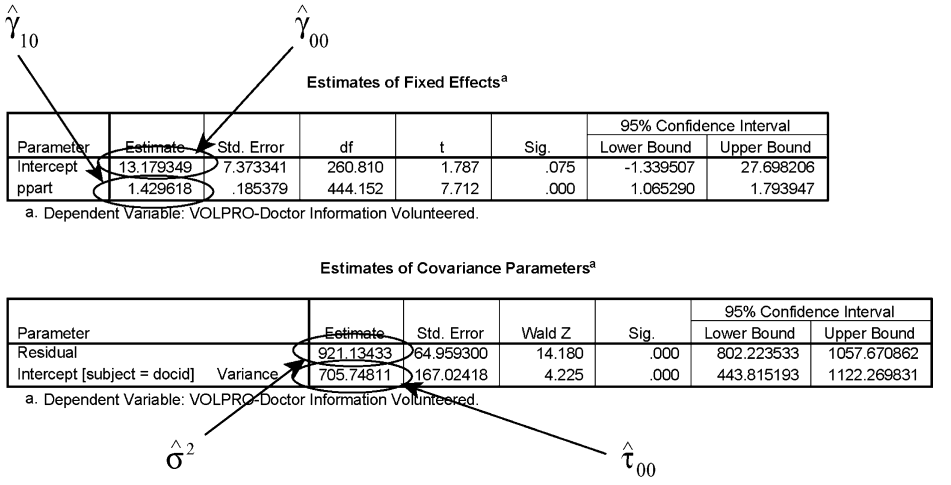


Figure 3 SPSS MIXED output for Model 2.

the deviance of this model (Model 2). The deviance of this model is 4,446.808, which is a reduction of 4,502.314 − 4,446.808 = 55.506 from the deviance of the equivalent model without the patient participation effect estimated (Model 1). Models 1 and 2 differ by only a single parameter estimate ($\hat{\gamma}_{10}$), so the difference between these deviances is distributed as chi-square with a single degree of freedom: $\chi^2(1) = 55.506$, $p < .0001$.[6]

By comparing $\hat{\sigma}^2$ between Model 2 and Model 1, we can derive a statistic that is sensitive to how much of the variance remaining in the number of volunteered utterances unaccounted for by doctor differences is attributable to patient participation. The residual variance from Model 1 is 1,056.484. Adding patient participation as a level-1 predictor reduces the residual variance to 921.134 in Model 2. One minus the ratio of these two variance estimates is often interpreted as a "variance accounted for" measure. Compared to when there are no level-1 predictors in the model, the addition of patient participation to the level-1 model changes the residual variance by a factor of 921.134/1,056.484 = 0.872, which translates into a "variance accounted for" measure of 1 − 0.872 = 0.128. This is conceptually equivalent to a squared partial correlation in ordinary least squares regression because it benchmarks variance explained by patient participation relative to variance remaining after removing variance accounted for by differences between doctors in how much they volunteer. As such, it cannot be interpreted analogously to the incremental R^2 (or "change in R^2") in ordinary single-level regression, as the incremental R^2 benchmarks variance explained relative to *total variance* in the outcome.

Notice a substantial shift in $\hat{\gamma}_{00}$ from Model 1 to Model 2. In Model 1, we interpreted $\hat{\gamma}_{00}$ as the mean number of utterances volunteered across doctors. In Model 2, the interpretation of $\hat{\gamma}_{00}$ is conditioned on patient participation being set to zero. So we interpret $\hat{\gamma}_{00}$ as the mean number of utterances, calculated across

doctors, when interacting with patients who do not participate (i.e., PPART = 0) and u_{0j} as how much doctor j's conditional mean varies from this grand conditional mean. Considering that in no doctor–patient consultation you would expect a patient to be totally silent (unless of course the patient is simply unable to communicate), this interpretation of $\hat{\gamma}_{00}$, although literally and statistically correct, really does not make much sense. But the fact that $\hat{\gamma}_{00}$ is conditioned on PPART being zero means that so, too, is the interpretation of $\hat{\tau}_{00}$, as it quantifies the variance of the conditional mean number of utterances volunteered—conditioned on PPART being zero. However, in this case and as you will see, this variance estimate actually is unconditional in this model, in that it is independent of the value of PPART you condition the interpretation on. This is not always true, however.

In MLM, it is common to construct a transformation of the predictor variables to give the intercept parameters, $\hat{\gamma}_{00}$ and $\hat{\tau}_{00}$, more useful interpretations. One common form of centering is *grand mean centering*. A variable is grand-mean-centered by subtracting the sample mean from each case's score on the variable. For example, in the 450 patients in the data file, the average participation score is 32.333 utterances. Patient participation would be grand-mean-centered by subtracting 32.333 from every case in the data file (i.e., $\text{PPARTC}_{ij} = \text{PPART}_{ij} - 32.333$; see Appendix A for the SPSS command). Reestimating this model but substituting PPARTC_{ij} for PPART_{ij} in the model (Model 3 in Table 1) changes $\hat{\gamma}_{00}$ substantially (from 13.179 to 59.403) but $\hat{\gamma}_{10}$ and $\hat{\tau}_{00}$ not at all. So this recentering has not done anything to the relationship between patient participation and how much a doctor volunteers. But we can now interpret $\hat{\gamma}_{00}$ as the estimated mean number of volunteered utterances when doctors interact with patients who participate an average amount (because $\text{PPARTC}_{ij} = 0$ when PPART_{ij} is equal to the sample mean), and $\hat{\tau}_{00}$ is interpreted as the variance of the doctor means for such patients. The test of this variance component tests a hypothesis that is conceptually equivalent to a random effects analysis of covariance. It tests whether doctors differ in how much they volunteer after controlling for doctor-level differences in how much their patients participate during the consultation. The results in Table 1 indicate that even after controlling for between-doctor differences in how much their patients participate during the consultation, the doctors still differ from each other, on average, in how much information they volunteer to their patients, $Z = 4.225$, $p < .001$. The likelihood ratio test yields $\chi^2(1) = 156.735$, $p < .0001$, from 4,603.543 − 4,446.808, where 4,446.808 is the −2LL for Model 2 and 4,603.543 is the −2LL from the same model but fixing the intercept rather than allowing it to vary randomly across level-2 units (the latter deviance is not presented in Table 1). For a detailed discussion of centering and its effects on model interpretation, see Raudenbush and Bryk (2002, pp. 31–35), Kreft and de Leeuw (2002, pp. 106–114), and Kreft, de Leeuw, and Aiken (1995).

Controlling for an additional level-1 predictor

Using available data on the length of each doctor–patient interview, measured in minutes, the relationship between how much a patient participates and the length of

the interview is positive, $r = .46$. Furthermore, doctors do volunteer more information in lengthier interviews, $r = .31$. So it should not be too surprising that doctors volunteer more when their patients talk more during lengthier consultations. It would be interesting to know whether patient participation is related to information volunteered from the doctor independent of interview length. Rephrased, keeping consultation length constant, do patients who participate relatively more still prompt their doctors to volunteer more information? If so, this suggests that the volunteered information per unit of time increases as a function of participation, not just total time spent interacting.

We answer this question by adding a second variable to the level-1 model, the length of the consultation in minutes, and assess whether the positive relationship between patient participation and information volunteered remains. To render the intercept interpretable, we grand-mean-center the length of the interview ($\text{LENGTHC}_{ij} = \text{LENGTH}_{ij} - 20.724$; see Appendix A for the SPSS syntax to do this transformation), treating the effect of consultation length as fixed across doctors and also use the grand-mean-centered measure of patient participation. The model in multilevel equations form is

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}\text{PPARTC}_{ij} + \beta_{2j}\text{LENGTHC}_{ij} + r_{ij}$$

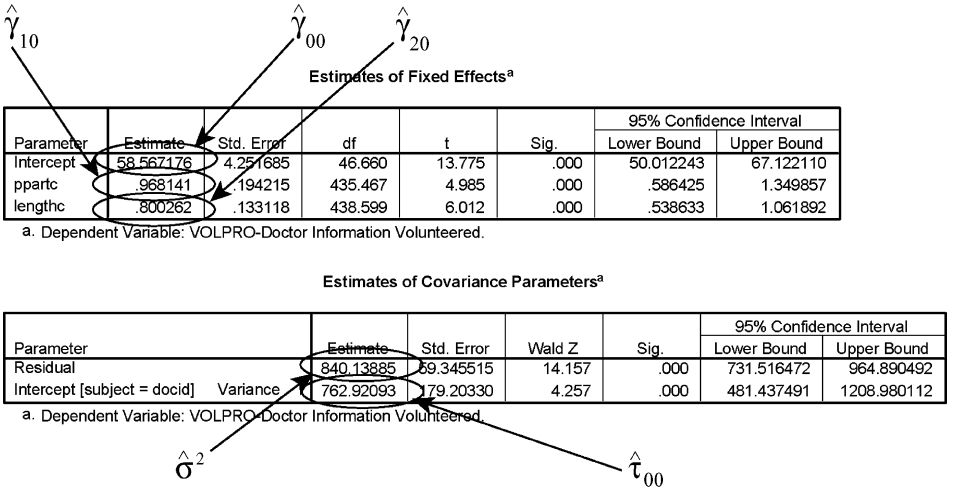
$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}; \quad \beta_{1j} = \gamma_{10}; \quad \beta_{2j} = \gamma_{20}$$

In mixed-model form, the model is

$$Y_{ij} = \gamma_{00} + \gamma_{10}\text{PPARTC}_{ij} + \gamma_{20}\text{LENGTHC}_{ij} + u_{0j} + r_{ij}$$

The SPSS output can be found in Figure 4 and Table 1 (Model 4). In this model, $\hat{\gamma}_{10} = \hat{\beta}_{1j} = 0.968$, which is positive and statistically different from zero, $t(435) = 4.985$, $p < .001$. Patients who participate relatively more solicit relatively more volunteered information from their doctors independent of the length of the consultation. Additionally, independent of patient participation, doctors volunteered relatively more information during relatively long interviews, $\hat{\gamma}_{20} = \hat{\beta}_{2j} = 0.800$, $t(439) = 6.012$, $p < .001$. Finally, among patients who participate an average amount during an interview of average length, the estimated mean number of units volunteered by a doctor is $\hat{\gamma}_{00} = 58.567$.

Comparing $\hat{\sigma}^2$ between models 4 and 1 gives us a means of quantifying the proportion of the variance remaining after partialing out between-doctor differences in information volunteered that can be explained by both patient participation and length of consultation. Using information from Table 1, this so-called "Level-1 pseudo R^2" is $1 - (840.139/1{,}056.484) = .205$. This is not literally an R^2 as in ordinary regression, however, as this proportion is indexed relative to variance in information volunteered *remaining after accounting for differences between doctors* rather than *total variance* in information volunteered. As such, it is closer to a squared partial correlation. However, it can be interpreted like an R^2 so long as one is clear that the measure is conditioned on the variance to be explained that is not already explained



Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	58.567176	4.251685	46.660	13.775	.000	50.012243	67.122110
ppartc	.968141	.194215	435.467	4.985	.000	.586425	1.349857
lengthc	.800262	.133118	438.599	6.012	.000	.538633	1.061892

a. Dependent Variable: VOLPRO-Doctor Information Volunteered.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		840.13885	59.345515	14.157	.000	731.516472	964.890492
Intercept [subject = docid]	Variance	762.92093	179.20330	4.257	.000	481.437491	1208.980112

a. Dependent Variable: VOLPRO-Doctor Information Volunteered.

Figure 4 SPSS MIXED output for Model 4.

by differences between the doctors, what is sometimes called "individual-level variance." It is possible to derive a measure of "incremental individual-level variance explained" by the addition of length of consultation by noting that, as derived earlier, patient participation by itself explained about 12.8% of the variance unexplained by differences between doctors. Adding length of consultation to the model increased the Level-1 R^2 from .128 to .205. Thus, we can say that length of consultation uniquely explains an additional 20.5 − 12.8 = 7.7% of the variance in how much information a patient volunteers after accounting for differences between doctors, but now controlling for how much the patient participates. A similar reasoning would be used to derive the proportional increase in variance accounted by patient participation independent of length of consultation by reversing the order of entry of these two variables. For more details on quantifying variance explained at various levels, see Hox (2002, pp. 63–71) and Snijders and Bosker (1999, pp. 99–109).

Testing for between-doctor differences in the relationship between participation and information volunteered

Models 3 and 4 fix the effect of participation to be constant across doctors. Such a constraint might be unrealistic. Perhaps, physicians differ with respect to the relationship between participation and information volunteered. Some doctors might be quite responsive to such participation, whereas others are not. In ordinary regression, variation in a regression coefficient is typically modeled by including an interaction between the predictor variable of interest and some second variable that the researcher anticipates or hypothesizes is related to that variable's effect—the moderator variable (see, e.g., Aiken & West, 1991; Hayes, 2005). Unfortunately, this approach requires that the researcher have an explicit model of how a variable's effect is influenced or changed by a second variable and that variable must be measured in

order to assess its role as a moderator. MLM allows the investigator to ask simply whether the effect of a level-1 variable varies across level-2 units without imposing the requirement, as moderated multiple regression does, that some second variable be proposed as the moderator variable. This question is answered by setting a level-1 variable's effect as random and assessing whether the variance of the random component is statistically different from zero. In this case, we will extend Model 4 by estimating the coefficient for participation in the level-1 model as a random effect. In multilevel equations form,

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}\text{PPART}C_{ij} + \beta_{2j}\text{LENGTH}C_{ij} + r_{ij}$$
$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}; \quad \beta_{1j} = \gamma_{10} + u_{1j}; \quad \beta_{2j} = \gamma_{20}$$

or, in mixed model form by substituting the level-2 model into the level-1 model,

$$Y_{ij} = \gamma_{00} + (\gamma_{10} + u_{1j})\text{PPART}C_{ij} + \gamma_{20}\text{LENGTH}C_{ij} + u_{0j} + r_{ij}$$

The u_{1j} in the level-2 model of β_{1j} is the random component of the effect of participation. Its presence in the model is what allows the effect of participation to vary across doctors. In this model, γ_{10} is the average effect of participation across doctors, and u_{1j} quantifies how the effect of participation for doctor j differs from this average.

The SPSS output can be found in Figure 5 and is summarized in Table 1 (Model 5). In this model, $\hat{\gamma}_{10} = 0.989$, which is positive and statistically different from zero, $t(41) = 3.333$, $p < .01$. On average, the relationship between patient participation and information volunteered is positive. Patients who participate relatively more tend to solicit more volunteered information from their doctor. Independent of patient participation, doctors volunteered more information during relatively longer interviews, $\hat{\gamma}_{20} = \hat{\beta}_{2j} = 0.864$, $t(406) = 6.319$, $p < .001$. Finally, among patients who participate an average amount during an interview of average length, the estimated mean number of units volunteered by a doctor is $\hat{\gamma}_{00} = 59.564$.

The part of the output central to question of interest can be found in the "Estimates of Covariance Parameters" section. Notice that this section has changed considerably from prior models as a result of the additional subcommand "COVTYPE (UN)" (see Appendix A), which stands for a *covariance structure of type unstructured*. When more that two effects are set as random, it is possible not only to estimate the variance of those effects but also their *covariance*, or relationship with each other. Recall that u_{0j} represents the random component of the intercept for doctor j—conceptually, the difference between the intercept for doctor j and the average intercept—and u_{1j} represents the random component of the participation effect for doctor j, which is conceptually the difference between the effect of patient participation for doctor j and the average participation effect. Usually, we would want to allow these random components to intercorrelate, reflecting the fact that doctors who volunteer relatively more (or less) than other doctors may be more (or

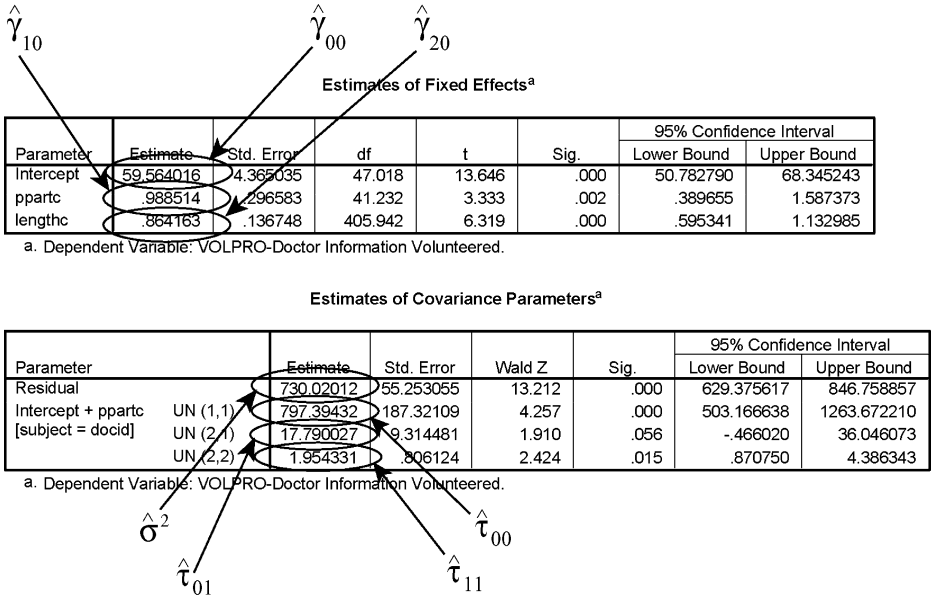


Figure 5 SPSS MIXED output for Model 5.

less) affected by patient participation. The unstructured covariance type allows this. Forcing these random effects to be uncorrelated would be unnatural in most circumstances, but if you wanted to do that, you could specify "vc" as the covariance type, which stands for "variance components."

Different programs will display the variances differently in their output. In SPSS, UN(1,1) is the variance of the random intercept components ($\hat{\tau}_{00}$), and UN(2,2) is the variance of the random components of the participation effect ($\hat{\tau}_{11}$). UN(2,1) is the covariance between these random components, often denoted $\hat{\tau}_{01}$. As can be seen from the output, $\hat{\tau}_{00} = 797.469$ and is statistically different from zero, interpreted as significant between-doctor variation in how much doctors volunteer, on average, with a patient who participates an average amount during a consultation of average length. Furthermore, the variance across doctors in the effect of participation is $\hat{\tau}_{11} = 1.954$ and statistically different from zero. So it seems that the effect of patient participation varies across doctors. The interpretation of $\hat{\tau}_{01}$ is facilitated by converting it to a correlation. This is accomplished by dividing it by the square root of the product of $\hat{\tau}_{00}$ and $\hat{\tau}_{11}$. From Figure 5,

$$r_{01} = \frac{\hat{\tau}_{01}}{\sqrt{\hat{\tau}_{00}\hat{\tau}_{11}}} = \frac{17.790}{\sqrt{797.469 \times 1.954}} = 0.451$$

This is the estimated correlation between the random components and can be interpreted to mean that the coefficient for patient participation is larger among doctors who tend to volunteer relatively more information on average. But by a Wald test, this correlation is not statistically different from zero.

The alternative test of the random effect of patient participation, the likelihood ratio test, is calculated by comparing the deviances of models 5 and 4. Recall that Model 5 is identical to Model 4 except that the effect of patient participation in Model 4 is fixed. For Model 4, −2LL is equal to 4,412.499, whereas −2LL equals 4,389.625 for Model 5, a difference of 4,412.499 − 4,389.625 = 22.874. Models 5 and 4 differ by two parameter estimates ($\hat{\tau}_{11}$ and $\hat{\tau}_{01}$), so $\chi^2(2) = 22.874$, $p < .001$. But this p value tests the null hypothesis that the random effect of patient participation and the covariance between this random effect and the random intercept are *both* zero. To get a p value for just the random effect of patient participation, cut this p value in half (Snijders & Bosker, 1999, pp. 90–91). This is a situation where the likelihood-ratio test and the Wald test conflict. If you happen to conduct both and get conflicting results, trust the likelihood ratio test more.

Modeling the between-doctor differences in the effect of participation

The prior analysis provided evidence that there is between-doctor heterogeneity in the relationship between patient participation and how much information a doctor volunteers during the consultation. But it leaves unanswered what, if anything, accounts for that heterogeneity. In this final example of MLM, I ascertain whether any of the heterogeneity is accounted for by how busy the doctor is. Busy doctors may have less free time to dedicate to patients who participate more, whereas doctors who can work at a more leisurely pace may be able to afford the extra time that patient participation requires, resulting in a greater flow of information between doctor and patient. Here, a doctor's busyness is operationalized as the average number of patients seen per day. Doctors who see more patients per day, all other things being equal, have less time to spend with each patient. Although this is not a perfect operationalization of the construct (as we do not know, for instance, whether doctors who see many patients work longer hours), it will suffice for this example. The variable name in the data is BUSY, and it is a level-2 variable as it is a property of the doctor rather than the patient.

In Model 5, β_{1j} quantifies the effect of participation on information volunteered for doctor j. To assess whether this coefficient varies as a function of how busy a doctor is, a multilevel model is estimated in which BUSY is used a predictor of *both* β_{1j} and β_{0j}. In multilevel equations form,

$$\text{Level 1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}\text{PPARTC}_{ij} + \beta_{2j}\text{LENGTHC}_{ij} + r_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}\text{BUSY}_j + u_{0j}; \quad \beta_{1j} = \gamma_{10} + \gamma_{11}\text{BUSY}_j + u_{1j}; \quad \beta_{2j} = \gamma_{20}$$

where BUSY_j is doctor j's average number of patients per day. As always, the mixed model form is created by substituting the level-2 equations into the level-1 equation,

$$Y_{ij} = \gamma_{00} + \gamma_{01}\text{BUSY}_j + (\gamma_{10} + \gamma_{11}\text{BUSY}_j + u_{1j})\text{PPARTC}_{ij} + \gamma_{20}\text{LENGTHC}_{ij} + u_{0j} + r_{ij}.$$

This form of the model makes it more transparent how γ_{11} quantifies the extent to which patient participation varies as a function of doctor busyness. In this representation of the model, the coefficient for patient participation, β_{1j}, is defined as $(\gamma_{10}+\gamma_{11}\text{BUSY}_j+u_{1j})$. Clearly, if γ_{11} is positive, the effect of patient participation increases with increasing busyness, whereas if γ_{11} is negative, the effect of patient participation decreases with increasing busyness. If γ_{11} is zero (or at least not statistically different from zero), then doctor busyness is unrelated to the effect of participation. This interpretation of the effect of the size of γ_{11} on β_{1j} is justified because γ_{10} is constant in the model and the random component u_{1j} is assumed uncorrelated with busyness.

The previous equation can be rewritten equivalently as

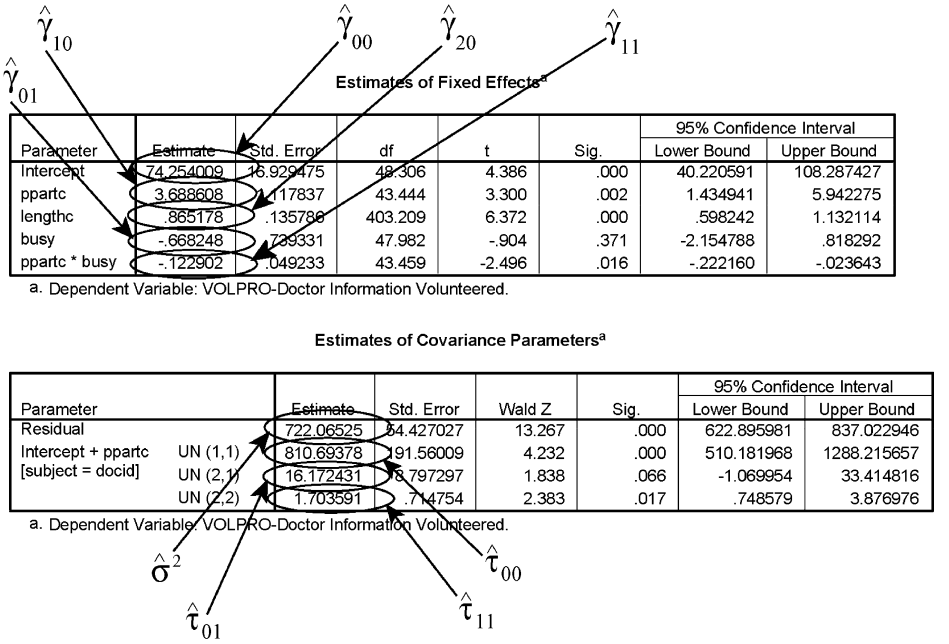
$$\begin{aligned} Y_{ij} = {} & \gamma_{00}+\gamma_{01}\text{BUSY}_j+(\gamma_{10}+u_{1j})\text{PPARTC}_{ij}+\gamma_{11}\text{BUSY}_j\text{PPARTC}_{ij} \\ & +\gamma_{20}\text{LENGTHC}_{ij}+u_{0j}+r_{ij} \end{aligned}$$

This form shows the similarity between this model and an ordinary regression model with interactions between the predictors. Notice that if you eliminate the two random components (u_{0j} and u_{1j}), this model reduces to a moderated multiple regression model.

The focus of the analysis is on the size of γ_{11}, which is the weight given to the product of how busy doctor j is with how much patient i participates during the consultation. Because BUSY, a level-2 variable, is postulated in this model to be related to the magnitude of a coefficient in the level-1 model, we say this model contains a *cross-level interaction*. The SPSS output corresponding to this model can be found in Figure 6 and Table 1 (Model 6). In this output, $\hat{\gamma}_{11}$ is found in the "Estimates for Fixed Effects" section of the output in the PPARTC $\times$ BUSY row. As can be seen, $\hat{\gamma}_{11} = -0.123$, $t(43) = -2.496$, $p = .016$. The coefficient is significantly negative. We interpret this to mean that (a) the effect of patient participation depends on how busy the doctor is and (b) the coefficient for patient participation is smaller among doctors who are more busy. Unfortunately, given the scaling of BUSY, we cannot be more precise than this. As we will illustrate, grand-mean-centering BUSY will allow us to be more specific about the nature of this cross-level interaction.

It is tempting to interpret several of the coefficients in this model as if they are "main effects." For instance, one might look at the output and conclude that after controlling for length of the consultation, doctor busyness, and the interaction between participation and busyness, the effect of participation is larger than when the cross-level interaction between busyness and participation was excluded. Similarly, you might interpret the nonsignificant coefficient of BUSY to mean that there is no relationship between how busy a doctor is and how much information he or she volunteers to a patient. Although this *may* be true, the coefficient for BUSY has a more complicated interpretation.

From the SPSS output, we have the following estimates in terms of the effects in the multilevel model represented above: $\hat{\gamma}_{00} = 74.254$, $\hat{\gamma}_{10} = 3.689$, $\hat{\gamma}_{20} = 0.865$, and



Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	74.254009	16.929475	49.306	4.386	.000	40.220591	108.287427
ppartc	3.688608	1.117837	43.444	3.300	.002	1.434941	5.942275
lengthc	.865178	.135786	403.209	6.372	.000	.598242	1.132114
busy	-.668248	.739331	47.982	-.904	.371	-2.154788	.818292
ppartc * busy	-.122902	.049233	43.459	-2.496	.016	-.222160	-.023643

a. Dependent Variable: VOLPRO-Doctor Information Volunteered.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		722.06525	54.427027	13.267	.000	622.895981	837.022946
Intercept + ppartc [subject = docid]	UN (1,1)	810.69378	191.56009	4.232	.000	510.181968	1288.215657
	UN (2,1)	16.172431	8.797297	1.838	.066	-1.069954	33.414816
	UN (2,2)	1.703591	.714754	2.383	.017	.748579	3.876976

a. Dependent Variable: VOLPRO-Doctor Information Volunteered.

Figure 6 SPSS MIXED output for Model 6.

$\hat{\gamma}_{01} = -0.668$. But only two of these have a useful substantive interpretation. We can legitimately say that doctors volunteer relatively more information during lengthier consultations, $\hat{\gamma}_{20} = 0.865$, $t(403) = 6.372$, $p < .001$, after controlling for the effects of patient participation, doctor busyness, and their interaction. Every other estimate in the output is *conditional* in one way or another. The coefficient for BUSY, $\hat{\gamma}_{01} = -0.668$, represents the relationship between doctor busyness and how much a doctor volunteers, on average, for a consultation of *average length* (i.e., conditioned on $\text{LENGTHC}_{ij} = 0$). The coefficient for PPARTC is $\hat{\gamma}_{10} = 3.689$. By a similar reasoning, this represents the estimated effect of patient participation conditioned on BUSY being zero. In other words, it can be interpreted as how patient participation is related to information volunteered among a group of doctors whose average number of patients per day is zero. Clearly, this is nonsensical. Finally, the estimate of the intercept, $\hat{\gamma}_{00} = 74.254$, quantifies the average information volunteered by doctors who have an average of no patients per day (i.e., when $\text{BUSY}_j = 0$) when spending an average amount of time with those patients (i.e., when $\text{LENGTHC}_{ij} = 0$). Again, this is nonsensical.

Such silly but statistically correct interpretations of some of these coefficients are produced by the scaling of BUSY and the fact that BUSY = 0 has no substantively sensible interpretation given the nature of the study. The solution to this is to make the zero point a location on the scale that has a sensible interpretation. Grand-mean-centering BUSY will accomplish this. The mean of BUSY in the data is 22.051. If we were to grand-mean-center BUSY by subtracting this average from each doctor's BUSY score (i.e., $\text{BUSYC}_j = \text{BUSY}_j - 22.051$) and then reestimate the model, this

rescaling of BUSY would produce coefficients that are all sensible and substantively interpretable. The coefficients from this model are presented in Table 1 (Model 7). As can be seen, grand-mean-centering BUSY has changed the intercept ($\hat{\gamma}_{00} = 59.581$) and the coefficient for patient participation ($\hat{\gamma}_{10} = 0.979$). Because $\text{BUSYC}_j = 0$ when BUSY_j is at the sample mean, the intercept can now be interpreted as the estimated mean amount of information volunteered for a doctor who is average on the number of patients he or she sees in a day when consulting with a patient for an average amount of time. The coefficient for patient participation is the effect of a one-unit difference in patient participation on how much a doctor volunteers when interacting with a patient and who is average in how many patients he or she sees in a day. This coefficient is positive and statistically different from zero.

With this estimate of the effect of participation when a doctor is average in terms of busyness, we can interpret the sign and size of the cross-level interaction between doctor busyness and patient participation more easily. This coefficient is negative and statistically different from zero, $\hat{\gamma}_{11} = -0.123$. It estimates the difference in the coefficient for patient participation between two doctors who differ by one patient per day in how busy they are. We already know for a doctor of average busyness, the coefficient for patient participation is estimated to be 0.979. For a doctor who is *one patient per day busier than average*, the coefficient for patient participation is $0.979 + (-0.123) = 0.856$. But for a doctor who is *one patient per day less busy than average*, the coefficient for patient participation is $0.979 - (-0.123) = 1.102$. The substantive interpretation is that patient participation has a bigger effect on how much a doctor volunteers among doctors who are less busy.

Multilevel modeling of repeated measures

Thus far, we have only considered multilevel models of an outcome variable in which the level-1 unit is a single individual, measured once on each variable, and who is only one of several individuals nested under a level-2 unit. But MLM is also useful when the individual participant in a study is the level-2 unit and measurements of the individual are taken repeatedly, either on the same variable or on several different variables. In that case, measurement occasion is the level-1 unit, which is nested under individual, the level-2 unit, and research questions and hypotheses typically focus on *change* in the measurements over repeated measurements as well as predictors of the extent of that change. Although such designs have historically been analyzed with repeated measures analysis or variance of equivalent methods, MLM is increasingly seen as the method of choice for the analysis of data from longitudinal or repeated measures designs. MLM makes it easy to answer questions such as "How much on average do individuals change over time"? "What is the rate of change"? "Do individuals differ with respect to how much they change"? and "What predicts how much or how quickly people change"?

Two recent examples of longitudinal MLM can be found in Slater, Henry, Swaim, and Anderson (2003) and Snyder, Fleming-Milici, Slater, Sun, and Strizhakova (2006). Slater et al. (2003) were interested in the age-old question as to whether

exposure to media violence prompts aggression. Four times over a 2-year period, teenagers were measured with respect to their exposure to various forms of violent media (e.g., television shows, video games, movies). Their aggressive thoughts were also measured at these same four times. By *group-mean-centering* (see, e.g., Kreft & de Leeuw, 1998, and Paccagnella, 2006, for a discussion of group- vs. grand-mean-centering and Singer & Willett, 2003, for a discussion in the context of longitudinal models) each teenager's measurement on exposure to violent media content, they assessed whether a teenager, when exposed to more violent media content than he or she usually is, tends to be more or less aggressive than he or she usually is. They also used time as a predictor to assess whether there were individual differences in how quickly a teenager's aggressive thinking changes over time and what predicts that rate of change, such as Internet use, age, and sensation seeking.

The multilevel analysis of longitudinal data carries with it all of the advantages of multilevel analysis while overcoming many of the shortcomings of repeated measures analysis of variance. For instance, there is no requirement that measurement begins and ends at the same time for every individual in the study, the measurement periods do not have to be equally spaced, the number of measurement periods can vary between individuals, it is relatively easy to test models of the factors that predict how much or how quickly a person changes, and models of change can be estimated that do not presume that change is linear over time. It can also be used to model long sequences of measurements collected dozens or even hundreds of times on the same person in different situations or in response to different stimuli, such as physiological responses like heart rate, electrodermal activity, or blood chemistry, without grouping the measurements into arbitrary categories of time (see, e.g., Hruschka, Kohrt, & Worthman, 2005, for a discussion on the use of MLM to analyze physiological data). Many of the concepts discussed already translate easily to the analysis of longitudinal data. For guidance on longitudinal MLM, see Byrne and Crombie (2003); Bryk and Raudenbush (1987); Duncan, Duncan, Strycker, Li, and Alpert (1999); Hox (2002, pp. 73–102); Raudenbush and Bryk (2002, pp. 160–204); Luke (2004, pp. 62–72); Snijders and Bosker (1999, pp. 166–199); Singer and Willett (2003); and Willett (1997). Singer (1998) and Peugh and Enders (2005) have good tutorials on estimating longitudinal multilevel models in SAS and SPSS, respectively.

In a longitudinal multilevel model, each level-2 unit is measured repeatedly on the same variable, and the focus of the analysis is on estimating change in the outcome variable over time and predictors of that change. MLM can also be used with repeated measures designs in which each level-2 unit is measured repeatedly on a variable, each time in response to one of several different stimuli or situations the unit is presented that vary between themselves in important measured ways. For example, a researcher might show each participant in a study (the level-2 unit) several print advertisements (the level-1 unit) that vary between themselves on one or more dimensions (e.g., the type of product advertised, the size of the advertisement, whether the print is color or black and white, the framing of the appeal in the ad, etc.). After each ad is presented, the researcher might measure such outcome

variables as how much the participant likes the ad or his or her intention to purchase the product. MLM would allow the investigator to build a model of the outcome variable as a function of features of the ads (level-1 variables) as well as characteristics of the participants such as their gender, age, personality, or other individual differences (level-2 variables). By estimating the level-1 variables as random effects, the investigator could assess the extent to which the impact of features of the advertisements varies between people. Cross-level interactions could be included to assess whether between-person heterogeneity in the impact of ad features depends systematically on one or more of the individual differences measured.

Conclusion and next steps

In any area of life, proficiency comes with patience and practice. MLM can seem overwhelming to the uninitiated. My hope is that with the information presented here, you will have begun to build the foundation needed to pursue some of the intricacies of MLM in greater detail. There are numerous excellent books on the market today, and articles that both discuss and use MLM are appearing with increasing frequency throughout the social sciences. I recommend *Multilevel Modeling* of Luke (2004) as a good next step; Cohen, Cohen, West, and Aiken (2003) provide two good chapters on MLM of cross-sectional and longitudinal data. On the other extreme, *Hierarchical Linear Models* of Raudenbush and Bryk (2002) should eventually be in your library, as no other book gives both applied advice and tremendous technical detail for those so inclined. Between these endpoints on the applied–to–technical continuum, you cannot go wrong with Hox (2002), Kreft and de Leeuw (1998), and Snijders and Bosker (1999). Although this primer focused on the use of SPSS, there are numerous programs you may find better suited to the needs of your specific area of research. For guidance on making the choice, see the set of reviews compiled online at the Center for Multilevel Modeling at http://www.mlwin.com/.

Notes

1 I thank Don Cegala for providing the data set used as the foundation for the analyses reported here. For pedagogical purposes, the data from the study were modified, including artificially increasing the sample size and the construction of a one variable not actually measured in the study, so the results of all analyses reported here should not be cited to support any claims made about the relationships between the variables discussed in this article. For details on the findings of this study, see Cegala et al. (in press).

2 Programs vary in how the data file must be set up for multilevel modeling. HLM, for example, uses a different file format than SPSS. For details on how to set up a multilevel data file for analysis in HLM, see the HLM manual.

3 SPSS uses a Satterthwaite approximation for the degrees of freedom, which typically yields a noninteger *df*. We report degrees of freedom to the nearest integer.

4 Some argue that this test should only be conducted using the deviances from models estimated using REML rather than regular maximum likelihood. Although the test does

tend to perform better when based on REML deviances, the difference is negligible and usually will not matter in terms of interpretation (Morell, 1998).

5 The convention in multilevel modeling is to analyze and interpret results from a multilevel model using data in unstandardized form.

6 In this case, because these models differ only by the inclusion of a fixed effect, the deviances for each model should be derived with regular maximum likelihood estimation and not restricted maximum likelihood (see Snijders & Bosker, 1999, p. 57).

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression*: *Testing and interpretation interactions*. Thousand Oaks, CA: Sage.

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they are independent. *Organizational Research Methods*, **7**, 400–417.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, **101**, 147–158.

Byrne, B., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, **2**, 177–203.

Cegala, D. J., Street, R. L., Jr., & Clinch, C. R. (in press). The impact of patient participation on physicians' information provision during a primary care medical interview. *Health Communication*.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling*. Mahwah, NJ: Lawrence Erlbaum.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, **26**, 441–462.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Lawrence Erlbaum.

Hox, J. (2002). *Multilevel analysis*. Mahwah, NJ: Lawrence Erlbaum.

Hruschka, D. J., Kohrt, B. A., & Worthman, C. M. (2005). Estimating between- and within-individual variation in cortisol levels using multilevel modeling. *Psychoneuroendocrinology*, **30**, 698–714.

Jackson, S., & Brashers, D. E. (1994). *Random factors in ANOVA*. Thousand Oaks, CA: Sage.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, **99**, 422–431.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, **30**, 1–22.

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

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, **25**, 271–284.

Morell, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, **54**, 1560–1568.
Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of group mean and the assessment of group effects. *Evaluation Review*, **30**, 66–85.
Pan, Z., & McLeod, J. M. (1991). Multilevel analysis in mass communication research. *Communication Research*, **18**, 140–173.
Peugh, J. L., & Enders, C. K. (2005). Using the SPSS MIXED procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, **65**, 717–741.
Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, **2**, 173–185.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
Ritchie, L. D., & Price, V. (1991). Of matters micro and macro: Special issues for communication research. *Communication Research*, **18**, 133–139.
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, **24**, 323–355.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Slater, M. D., Henry, K. L., Swaim, R. C., & Anderson, L. L. (2003). Violent media content and aggressiveness in adolescents: A downward spiral model. *Communication Research*, **30**, 713–736.
Snijders, T., & Bosker, R. (1999). *Multilevel analysis*: *An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
Snyder, L., Fleming-Milici, F., Slater, M. D., Sun, H., & Strizhakova, Y. (2006). Effects of alcohol advertising exposure on youth drinking. *Archives of Pediatric and Adolescent Medicine*, **160**, 18–24.
Willett, J. B. (1997). Measuring change: What individual growth modeling buys you. In E. Amsel & K. A. Rinniger (Eds.), *Change and development: Issues of theory, method, and application* (pp. 213–243). Mahwah, NJ: Lawrence Erlbaum.

Appendix A

SPSS Syntax to Estimate the Multilevel Models Described in this Paper

```
Model 1
  MIXED volpro/PRINT solution testcov/METHOD = ML
    /FIXED = intercept/RANDOM = intercept | SUBJECT(docid).
Model 2
  MIXED volpro WITH ppart/PRINT solution testcov/METHOD = ML
    /FIXED = intercept ppart/RANDOM = intercept | SUBJECT(docid).
Model 3
  COMPUTE ppartc = ppart − 32.333.
  MIXED volpro WITH ppartc/PRINT solution testcov/METHOD = ML
    /FIXED = intercept/RANDOM = intercept | SUBJECT(docid).
```

(continued)

Appendix A. *Continued*

```
Model 4
  COMPUTE ppartc = ppart − 32.333.
  COMPUTE lengthc = length − 20.724.
  MIXED volpro WITH ppartc lengthc/PRINT solution testcov/METHOD = ML
    /FIXED = intercept ppartc lengthc/RANDOM = intercept | SUBJECT(docid).
Model 5
  COMPUTE ppartc = ppart − 32.333.
  COMPUTE lengthc = length − 20.724.
  MIXED volpro WITH ppartc lengthc/PRINT solution testcov/METHOD = ML
    /FIXED = intercept ppartc lengthc
    /RANDOM = intercept ppartc | SUBJECT(docid) COVTYPE (un).
Model 6
  COMPUTE ppartc = ppart − 32.333.
  COMPUTE lengthc = length − 20.724.
  MIXED volpro WITH ppartc lengthc busy/PRINT solution testcov/METHOD = ML
    /FIXED = intercept ppartc lengthc busy busy * ppartc
    /RANDOM = intercept ppartc | SUBJECT(docid) COVTYPE (un).
Model 7
  COMPUTE ppartc = ppart − 32.333.
  COMPUTE lengthc = length − 20.724.
  COMPUTE busyc = busy − 22.051.
  MIXED volpro WITH ppartc lengthc busyc/PRINT solution testcov/method = ML
    /FIXED = intercept ppartc lengthc busyc busyc * ppartc
    /RANDOM = intercept ppartc | SUBJECT(docid) COVTYPE (un).
```