Communications of the Association for Information Systems

Volume 25 Article 45

12-2009

Assessing Information Technology Use over Time with Growth Modeling and Hierarchical Linear Modeling: A Tutorial

Robert F. Otondo
Mississippi State University, rotondo@business.msstate.edu

Tim Barnett
Mississippi State University

Franz W. Kellermanns
Mississippi State University

Allison W. Pearson
Mississippi State University

Rodney A. Pearson
Mississippi State University

Communications of the Association for Information Systems

CAIS

Assessing Information Technology Use over Time with Growth Modeling and Hierarchical Linear Modeling: A Tutorial

Robert F. Otondo

Tim Barnett

Franz W. Kellermanns

Allison W. Pearson

Rodney A. Pearson

Management & Information Systems Dept., Mississippi State University

rotondo@cobilan.msstate.edu

Abstract:

Time is an important factor in the use of information technology. However, traditional information systems research methods cannot adequately account for the dynamic nature of time-based relationships often found in longitudinal data. This shortcoming is problematic when investigating volatile relationships that evolve over time (e.g., information technology use across users, departments, and organizations). Educational, sociological, and management researchers study the influence of time using a rigorous multilevel method called *growth modeling*. We demonstrate the use of growth modeling in this tutorial, which is based on a semester-long study of an actual web-based university-level course content delivery system. The tutorial provides guidance on preliminary data tests, the construction and analysis of growth models using hierarchical linear modeling, and the interpretation of final results. The tutorial also describes other unique advantages of using growth modeling for IS research.

Keywords: IT use, individual use, individual behaviors, longitudinal research, tutorial.

Volume 25, Article 45, pp. 607-640, December 2009

The manuscript was received 3/2/2008 and was with the authors 12 months for 2 revisions.

Communications of the Association for Information Systems

I. INTRODUCTION

The consideration of temporal issues, particularly those that relate to how individuals, groups, and organizations *relate to time* and *act over time*, is an important research topic in both general organizational studies and information systems (IS) research [Ancona, Okhuysen, and Perlow 2001; Devaraj and Kohli 2003; Orlikowski and Iacono 2001; Petersen et al. 2002]. For example, time may be particularly important in information technology (IT) acceptance and adoption because individuals don't typically learn about, accept, adopt, or even reject information systems and technologies on just one instance or occasion [Orlikowski and Iacono 2001]. Moreover, an individual's perceptions of motivations toward IT can change over time, and the relationships between such changes and user intentions remain to be investigated [Malhotra et al. 2008]. As Benbasat and Barki [2007] argue:

> *Longitudinal studies that view and assess system use over time are likely to be particularly revealing, as they can help us better understand the fluid relationships that exist between an adoption model's constructs and a variety of mutually influential set of behaviors users typically engage in, such as their adaptation, learning, and hands-on usage behaviors, as well as the subsequent influence of these behaviors on users' future beliefs* (p. 215).

At the group or departmental level, many IS and IT issues, such as transaction processing, server response, network throughput, and web site browsing, are time-sensitive and are likely to affect group productivity, as well as group and customer satisfaction [Agarwal and Venkatesh 2002; Palmer 2002: 153]. Temporal factors related to the "timeliness" and "currency" of information also impact IS and IT decision-making at the organizational level, and such issues are relevant to the management of web site content and customer loyalty [Agarwal and Venkatesh 2002; Mithas et al. 2006–7]. Longitudinal studies from the organizational literature raise similar questions about changes in individual work in IS environments over time. Bliese and Ployhart [2002: 363] list a number of articles investigating temporal changes in social support, group consensus, and stressor-strain relationships [e.g., Bliese and Britt 2001], individual work performance and work performance criteria [e.g., Deadrick et al. 1997], and how individual abilities determine changes in work performance over time [e.g., Hofmann et al. 1993]. Kozlowski and Klein [2000] describe the potential impact of temporal changes in culture, work-flow interdependence, and task and budget cycles on individual actions over time. In turn, these studies can serve as blueprints for similar research in IS environments, such as changes in the characteristics of computer-enabled groups over time, changes in IT use, user performance, and user performance criteria, and how individual abilities, characteristics, and perceptions are associated with changes in IT use and IT-enabled work performance over time.

While the importance of temporal issues is widely acknowledged, for a variety of reasons the passage of time is difficult to incorporate into IS research. For example, IT use is sporadic and variable [Lee et al. 2006; Petersen et al. 2002]. This not only produces changing, dynamic patterns of IT use over time, but may also create complex interactions between IT use and important user characteristics such as gender, age, and competence. Further, the inadequate modeling of longitudinal data, characterized by a reliance on static or "snapshot" data for time-based research, produces a limited amount of data which inadequately describes complex, dynamic, and evolving human behaviors over time [Petersen et al. 2002: 74]. Finally, there are practical difficulties involved in time-based research, such as gaining access to the same organization members over time. Thus, a time-oriented, longitudinal approach has the potential to improve our ability to explain IT use in theoretical research and to offer practical advice to IT designers and stakeholders.

Successfully incorporating time into IT use research first requires that the history of each individual's use of IT must be described and analyzed. More problematically, it also requires that variation in the "within-person" descriptions of IT use must be related to variations in "between-person" characteristics [Bliese and Ployhart 2002; Raudenbush 2001]. Completing the first task without the second produces impoverished descriptions of events that ignore important elements of human nature and behavior. Completing the second task is difficult because the constructs and variables arise from two different levels of analysis: "event-level" variables describe incidents that vary significantly across time, while "person-level" variables describe human characteristics, such as race, gender, and trait differences that either do not change, or change little, across relevant time frames.

IS researchers often skirt such complexities by focusing on only one level of analysis, even though theory suggests the existence of *cross-level* effects between variables at different levels of analysis (e.g., how a person-level characteristic such as gender might be related to an event-level characteristic such as IT use). The study of cross-level effects became possible with the development of rigorous multilevel statistical methods for simultaneously analyzing the relationships among variables drawn from different levels of analysis. These methods go by many names, including *hierarchical linear modeling* (HLM) and *random coefficients modeling* (RCM). *Growth modeling* is a particular form of HLM founded on an event-level of analysis over time. The increasing use of growth modeling and other multilevel methods in management and education research suggests the potential usefulness of such methods in a wide variety of IS research involving time [e.g., Bryk et al. 1993; Klein and Kozlowski 2000; Bliese and Ployhart 2002; Raudenbush and Bryk 2002; Singer and Willett 2003; Hitt et al. 2007]. While the number of multilevel studies in IS research is growing [e.g., Ang et al. 2002; Burton-Jones and Straub 2005; Lapointe and Rivard 2005; Mithas et al. 2006–7], one finds very few using growth modeling methods [e.g., He et al. 2007]. Growth modeling could be applicable to a number of IT contexts in which variables and their interrelationships change over time. For example, one might develop a research question similar to the following generalized form: *To what extent does the relationship between Y (e.g., a dependent variable such as perceived usefulness or actual use) and X (e.g., an individual- or group-level variable such as gender, playfulness, or group cohesiveness) vary across time?*

Therefore, the purpose of this paper is to provide a tutorial that describes and explains the use (and usefulness) of growth modeling in IS research, utilizing a specific research model of IT acceptance and use. We will focus on HLM, though other applicable methods, such as structural equation modeling (SEM), can also be used. Though SEM-based growth modeling has some advantages (e.g., handling measurement error), an HLM approach is often preferred when "each case is observed at different time points, if the repeated measure is a count variable, or if there are three or more levels of analysis;" in addition, HLM is typically easier to use [Bollen and Curran 2006:54].

The tutorial is organized as follows. First, we provide a theoretical model and hypotheses for the purposes of demonstrating growth modeling. As such, our model is a basic model of system use, predicted by commonly associated demographic variables such as gender, age, and academic performance. This model is intentionally parsimonious. It is not a test of new theory, but is provided merely for illustrative purposes. We then explain growth modeling theory and methods, beginning with a discussion of how our tests of HLM assumptions affect the growth models used in our study. Next, we demonstrate growth modeling and analysis using the popular multilevel statistical package *HLM for Windows, version 6.06* [Raudenbush et al. 2008]. We describe many of our findings concerning relationships among time, IT use, and person-level characteristics, and draw conclusions about the applicability, usefulness, and future potential of growth modeling in IS research.

II. GROWTH MODELING ANALYSIS OF IT USE OVER TIME: AN EXAMPLE

The Level-1 Model: Information Technology Use over Time

The context for our tutorial is the study of a real web-based curriculum content delivery system, called *BIStro* (a fictionalized name), which was designed, developed, and implemented in a medium-sized university in the southeastern United States. The study evolved from concerns of BIStro's designers that their system might not serve all student populations equally. The designers based their concerns on several studies in the IS literature concerning the effects of user characteristics on IT use [e.g., Gefen and Straub 1997; Venkatesh and Morris 2000; Venkatesh et al. 2003; Hourcade et al. 2004; Ahuja and Thatcher 2005]. The system designers recognized their concerns involved relationships among *two* levels of analysis; (1) how person-level characteristics such as gender, age, and academic performance were associated with (2) event-level constructs such as IT use and time. We formalized our examination of students' use of BIStro with the following research question: *To what extent is information technology use over time related to gender, age, and academic performance?*

In developing the models for our study and tutorial, we conceptualized the role of time in IT use by drawing upon commonly-held demographic predictors of IT use, including age, gender, and academic performance. We deduced that BIStro users' experience (time) with a system was also important because users would be expected to learn and/or change their patterns of use over time based on increased familiarity with the system and on feedback from previous activities. We also expected that BIStro users might draw upon previous classroom experiences, such as increasing their use of the BIStro system shortly before exams [Brotherton and Abowd 2004]. Thus, our first hypothesis focuses on the relationship between event-level constructs of time and BIStro use.

HYPOTHESIS 1: *Web site usage will vary over time.*

The Level-2 Model: Academic Performance, Age, and Gender

Prior IS research supports the BIStro designers' concerns about person-level characteristics (e.g., gender, age, and academic performance) by showing important connections between these user characteristics and computer use. Research in the early-to-mid 1980s found that women and girls used home computers "less often and less intensively" than males [Papadakis 2000:2]. These findings are consistent with later research showing that the number of freshman males expressing interest in computer science is nine times that of freshman females [Malcolm et al. 2005]. Gender has also been associated with e-mail and Internet use [Fallows 2005; Gefen and Straub 1997].

Gender, as well as age and academic performance, can have direct and/or moderating effects on IT use over time. We will address the measurement of direct effects first. Growth modeling measures direct effects at the level-1 intercept (i.e., where *Time* = 0). Hofmann [1997, 2006] and Raudenbush and Bryk [2002: 150] recommend measuring direct level-2 effects after controlling for the level-1 independent variables. Following their advice, we composed the following hypothesis:

HYPOTHESIS 2a: *Gender directly relates to web site usage after controlling for time*.

We composed the next hypothesis to address gender's moderating level-2 effects on the level-1 equation:

HYPOTHESIS 2b: *Gender moderates web site usage over time*.

Research shows that home computer availability and use varies with age, increasing as people mature from their late twenties to their early forties, and then declining as they grow older [Burton-Jones and Straub 2005]. A report from the US Department of Commerce shows Internet usage rates remain relatively constant across college students aged 18–24 (70.6 percent) and working individuals aged 25–49 (i.e., 70.6 percent and 71.7 percent, respectively), though it drops to 64.6 percent in working individuals over 50 [US Dept. of Commerce 2004]. As such, age is a common predictor of system use.

HYPOTHESIS 3a: *User age directly relates to web site usage after controlling for time*.
HYPOTHESIS 3b: *User age moderates web site usage over time*.

Academic performance may also influence or moderate Internet use through the college years. A few studies have shown that academic performance is positively associated with home computer use, though by and large few studies examined the relationship between those variables [Subrahmanyam et al. 2000]. In our study, academic performance is operationalized as GPA at the start of the semester.

HYPOTHESIS 4a: *Previous academic performance directly relates to web site usage after controlling for time*.
HYPOTHESIS 4b: *Previous academic performance moderates web site usage over time*.

Research Model

We operationalized the dependent variable for the model, *BIStroUse*, as each participant's total number of log-ins to BIStro for each week of the study. We originally conceived of the event-level (level-1) independent variable, *Time*, as the length of time in weeks from the first assignment. However, we changed this definition of time somewhat after preliminary data analyses detected nonlinear patterns of BIStro use over the semester. Subsequent changes to the treatment of time will be explained later in the *Data Testing* section. Person-level (level-2) variables included gender, age, and the student's GPA at the beginning of the semester. Figure 1 depicts our research model.

III. GROWTH MODELING AND MULTILEVEL RESEARCH

Growth modeling offers a rigorous means of conceptualizing and analyzing IT use over time because it nests and integrates models simultaneously across levels of analysis. This "across level" architecture arises from multilevel theory's foundation in general system theory [von Bertalanffy 1968]. General systems theory explains how entities interact to form complex dynamic systems, such as when plants and animals interact to form ecosystems. Multilevel theory follows this line of thought by depicting team-, group-, organizational-, and other high-level phenomena as emerging from the interactions of individuals and/or lower-level subgroups [Kozlowski and Klein 2000].

While most multilevel studies nest individuals within teams, groups, and organizations, growth modeling "involves looking at how individuals (or units, groups, organizations, etc.) change over time and whether there are differences in patterns of change" [Bliese and Ployhart 2002: 363]. An excellent example of growth modeling can be found in Singer and Willett's analysis of Murnane et al.'s research on the labor-market experiences of 14–17 year old males [Murnane et al. 1999; Singer and Willett 2003]. Singer and Willett took results from a series of interviews about current employment (i.e., a series of event-level outcomes) and nested them within interviewees (i.e., within a

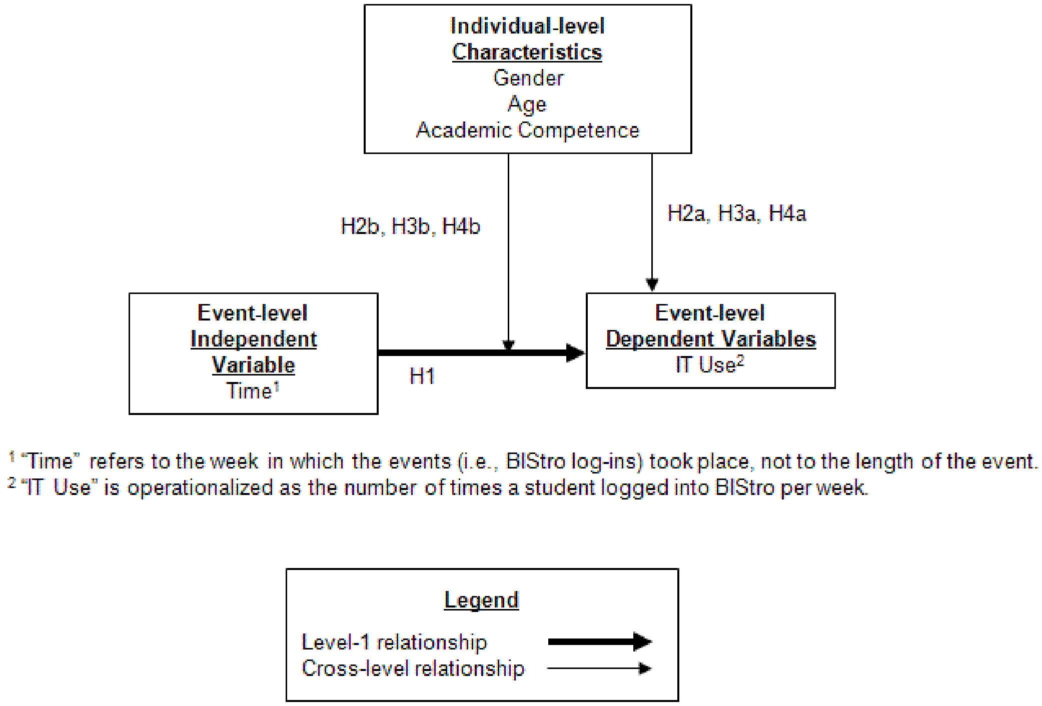


Figure 1. Research Model

higher-level entity—the individual subject). The study then associated the patterns of change across those events over time with person-level characteristics such as ethnicity and high school graduation. Nesting event-level phenomena within individuals is not typical of organizational multilevel research, which ordinarily nests person-level data within groups, teams, or organizational units. While different, growth modeling is nonetheless consistent with multilevel theory and methods [Raudenbush 2001; Singer and Willett 2003].

IS researchers can similarly investigate the effects of time [e.g., He et al. 2007]. In our study, the BIStro designers wanted to know the extent to which students' use of the BIStro system varied over time, and the extent to which that variation might be associated with important student characteristics such as gender, age, and academic performance. As in the Singer and Willet [2003] example, we nested an event-level sub-model within a higher-order, person-level model. We first constructed the event-level sub-model, with IT use as a dependent variable and time as the independent variable. We then nested that event-level sub-model into a person-level model, which allowed us to describe and later test both direct and moderating effects. Direct effects involved the influence of a student's personal characteristics on the event-level outcome (i.e., IT use), while moderating effects involved the influence of a student's personal characteristics on the relationship between time and IT use.

Growth Modeling *versus* Traditional IS Statistical Methods

Growth modeling's statistical techniques differ from those used in traditional IS research. For example, analysis of variance (ANOVA) is of limited use in growth modeling because its statistics about total variation (i.e., η^2) and between-group and within-group mean squares cannot parse variation between multiple levels of analysis. That is, ANOVA cannot distinguish what portion of total variation is accounted for by person-level variables, and what portion is accounted for by event-level variables. Accordingly, ANOVA has limited value in helping researchers understand cross-level effects.

Ordinary Least Squares (OLS) regression is also poorly suited for growth modeling research for at least three reasons. First, it cannot simultaneously incorporate variables from multiple levels of analysis. The incorporation of higher-level variables into an OLS regression would require that their effects are constant across lower-level variables—which is often not the case [James and Williams 2000]. Second, OLS regression requires random errors to be independent, normally distributed, and exhibit constant variance. In most cases, the inclusion of multiple groups would compromise these assumptions because the random errors are more likely to be similar within groups than across groups. Third, OLS regression cannot identify differences across a large number of groups and meaningfully quantify those differences within its structure (i.e., as intercepts and coefficients [Raudenbush and Bryk 2002]). Analysis of covariance (ANCOVA) suffers the same problems that limit OLS regression.

Structural equation modeling (SEM) is useful in some types of multilevel analysis, particularly when the multilevel model includes latent psychological constructs. This application of SEM in multilevel time analysis, called *latent growth analysis* (LGM), is advantageous because it accounts for measurement error. However, SEM analysis of longitudinal data requires that such data be "time-structured," that is, the data must be consistently collected across regular time intervals for all participants [Singer and Willett 2003]. Raudenbush and Bryk [2002: 187] note that this "forced choice" between LGM and HLM growth modeling techniques "reflects limitations in current software capabilities rather than limitations in modeling possibilities." They add that the choice between LGM and HLM growth modeling is based on the structure of the data; that is, (1) Are observed data balanced? (2) If "complete data" are balanced, are there missing data across time, and (3) Are complete data unbalanced? [Raudenbush and Bryk 2002: 186–199]. HLM growth modeling was deemed appropriate for our study because our data had few missing values and our predictors (i.e., gender, age, and GPA) can be reasonably assumed to be measured without error.

The limitations of ANOVA, OLS regression, ANCOVA, and SEM methods for analyzing longitudinal and multilevel data have led to the development of multilevel statistical methods such as HLM, which will now be briefly described. Those seeking fuller, more technical descriptions are encouraged to read the articles and books we cite in our tutorial, such as Bliese and Ployhart [2002]; Raudenbush and Bryk [2002]; and Singer and Willett [2003: ch. 4]. Texts by Klein and Kozlowski [2000] and Singer and Willett [2003] are recommended for those seeking deeper theoretical background. We also encourage interested readers to explore professional development workshops on multilevel modeling [e.g., Hofmann 2006]. Our tutorial uses the *HLM for Windows v6.06* software package, though our cited works sometimes use other software (e.g., Singer and Willett 2003 use SAS).

Growth Modeling with *HLM for Windows*

While other statistical packages also support random coefficient modeling, perhaps the easiest and most popular is *HLM for Windows* [Raudenbush et al. 2008]. We will describe how *HLM for Windows* can be used, but will not go into full detail given the space limitations for this article. Instead, readers are encouraged to learn the basics of *HLM for Windows* operation by using the above cited references and other publicly available resources. The latter includes web-based instruction manuals [e.g., Scientific Software International 2008; Raudenbush et al. 2004] and helpdesk web sites sponsored by university statistics departments [e.g., The University of Texas at Austin 2008]. These works can then be supplemented by our tutorial, which will focus on differences for growth modeling and issues of concern.

Simply put, random coefficient modeling describes multilevel models with *multiple* sets of regression equations. For example, a two-level random coefficient model would require two sets of regression equations. The first set contains one regression equation modeling a linear relationship between a dependent and one or more independent variables. This "level-1" equation is similar in many ways to OLS multiple regression. Both contain an intercept, one or more coefficients and variables, and an error term, but the level-1 multilevel equation differs because it must represent variables from two levels of analysis. An example of a level-1 equation from a two-level multilevel model is displayed in Equation 1 [Raudenbush and Bryk 2002: 100].

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \cdots + \beta_{Qj} X_{Qij} + r_{ij} \tag{1}$$

Each dependent and independent variable in Equation 1 (i.e., Y and X, respectively) contains at least two subscripts. The subscripts i and j describe how level-1 entities are nested within level-2 entities, such as when the i^{th} person is nested within the j^{th} school. While the dependent variable on the left side of the equation requires only two subscripts, the independent variables on the right side of the equation require an additional identification subscript, such as 1, 2, …, Q, to denote the presence of multiple variables, just as in OLS multiple regression.

Parameters on the right side of the equation (i.e., β) do not require the level-1 subscript, i, because they describe relationships between sets of i-level entities and their corresponding j-level entity (e.g., person i who is a member of group j). These relationships are described in terms of an intercept parameter and one or more coefficient or "slope" parameters. In addition to the j subscript, each parameter also needs an identification subscript which matches that of its corresponding variable (e.g., the Q^{th} coefficient and variable are denoted by β_{Qj} and X_{Qij}, respectively). As in OLS multiple regression, the first subscript "0" (e.g., as in β_{0j}) marks an intercept. Finally, the term r_{ij} represents level-1 residual for i-level entities nested within j-level entities.

The second set of regression equations describes how "level-2" variables interact with relationships between level-1 dependent and independent variables. These "cross-level" relationships can describe *direct effects* on the level-1 intercept β_{0j} and *moderating or rate of change effects* on the level-1 coefficients β_{1j}, β_{2j}, …, β_{Qj}. Multilevel models represent direct and rate of change effects by using the level-1 intercepts and slopes as outcome variables for the level-2 regressions, respectively, so the *i* subscript is unneeded. The level-1 intercept β_{0j} and slopes β_{1j}, β_{2j}, …, β_{Qj} from Equation 1 become outcome variables β_{qj} in Equation 2. In Equation 2, W_{qj} represents level-2 independent variables; γ_{qj}, the intercepts and slopes of the level-2 regressions; and u_{qj}, the level-2 error. The subscripts 0, 1, …, S_q identify the parameters and variables in the right side of the equation. The various level-2 equations can differ in the number of predictor variables they contain, so an additional subscript q is required (e.g., as in S_q). Equation 2 depicts the general form for level-2 regression equations [Raudenbush and Bryk 2002: 101].

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_{1j} + \gamma_{q2}W_{2j} + \cdots \gamma_{qS_q}W_{S_q j} + u_{qj} \quad (2)$$

Growth modeling, because it models person-level phenomena in level-2 instead of level-1 models, uses a slightly different nomenclature than other random coefficient models. The level-1 intercept β_{0j}, the slopes β_{1j}, β_{2j}, …, β_{Qj}, and the residual r from Equation 1 are moved to the level-2 model (i.e., Equation 4), with the group-level subscript j being changed to the person-level subscript i. The level-1 equation is often—but not always—much simpler because it typically contains one temporal variable (e.g., a person's age, the time of the event, or a number signifying one of several sampling "waves" [Singer and Willett 2003: 22]. The variable π is often used for level-1 parameters, while e is often used to represent the level-1 residual. The variable t is often used as a subscript for temporal, event-level aspects. A typical growth model is displayed in Equations 3 and 4.

$$Y_{ti} = \pi_{0i} + \pi_{1i}Time_{ti} + e_{ti} \quad (3)$$

$$\pi_{qi} = \beta_{q0} + \beta_{q1}X_{1i} + \beta_{q2}X_{2i} + \cdots \beta_{qS_q}X_{S_q i} + r_{qi} \quad (4)$$

IV. DATA ASSUMPTION TESTING

Before growth modeling can begin, the data must be tested to ensure its suitability for growth modeling analysis. Since HLM is "just regression" [Bickel 2007], the same assumptions still apply (e.g., homoscedasticity and linearity). As in regression, the lack of homoscedasticity can adversely affect the quality of statistical tests. These assumptions can be tested using traditional methods such as boxplots and scattergrams, and may be corrected with linear transformations. However, the use of multiple levels of analysis introduces additional assumptions as well. A list of key assumptions for two-level HLM and growth models can be found at Raudenbush and Bryk 2000: 255. These assumptions are listed in Appendix A, item 5.

As with other types of regression, assumption testing is an important part of HLM and growth modeling. We provide a list of key HLM assumptions and possible tests in the Appendix, item 5. However, given the space limitations for this tutorial, we refer the readers to more detailed explanations [e.g., Raudenbush and Bryk 2002: 252–287; Singer and Willett 2003: 127–132; and Snijders and Bosker 1999: 120–139]. We will also discuss results from those assumption tests that either indicated problems with our data and/or models, or may have been hidden in the various output files. First, we will provide a brief discussion of our data.

Data Entrainment

One of the first tasks in growth modeling is ensuring the data is properly *entrained*; that is, does the data present a history of IT use that matches "the rhythm, pacing, and synchronicity of processes that link different levels [of analysis]" [Kozlowski and Klein 2000: 24]? Our study met the entrainment requirement through the BIStro system's automatic collection and logging of event-level data. Participants entered a unique username and password when they logged into BIStro, thus creating sets of event-level (level-1) records describing BIStro activity that could be identified as originating from individual students. The ability to link event-level data to particular individuals allowed us to develop a corresponding person-level (level-2) data set describing each participant's gender, age, and academic performance. BIStro also timestamped each login, which allowed us to track the number of times a participant logged into BIStro each week across the semester. During each log-in, a student could access a number of BIStro functions, including quiz and homework assignment delivery as well as gradebook and calendar features. We collected data across four sections and two instructors, though the section and instructor data were not included in the study in order to simplify the model. We did not collect data during Week 7 of the study due to a network problem during that time. Data was stored in a person-period format, "in which each person has multiple records—one for each measurement occasion" [Singer and Willett 2003: 17]. The BIStro system was also able to provide a balanced data set that was largely free of missing data. Missing data is particularly troublesome when many subjects do not have sufficient data to provide sound individual-level regressions, so researchers should take appropriate steps beforehand to ensure unbiased and sufficient data are collected [e.g., Singer and Willett 2003: 157–159]. It also affects the type of analysis used in HLM, which we will discuss shortly.

Three hundred forty-seven students participated in the semester-long study. The participants logged into BIStro a total of 14,306 times during the sixteen weeks in which data was collected, an average of about 2.6 times per student per week. Other descriptive and correlation statistics are displayed in Table 1.

Table 1: Descriptive Statistics and Correlations [a]

Level	Variable	Mean	St. Dev.	Level 1 (1)	Level 1 (2)	Level 2 (3)	Level 2 (4)	Level 2 (5)
Event (Level-1)	(1) Week [b]	8.94	5.04			------	------	------
	(2) IT Use [c]	6.84	6.36	0.130***		------	------	------
Individual (Level-2)	(3) Gender [d]	0.58	0.49	------	------			
	(4) Age [e]	21.85	2.75	------	------	-0.026		
	(5) GPA	2.97	0.59	------	------	-0.144**	-0.245***	

[a] Reported correlations involve variables from the same level of analysis. Table cells containing "------" indicate "cross-level" relationships between variables from different levels of analysis, which are not amenable to correlation analysis.

[b] Week numbers run from 1 to 17; weeks 7 and 10 not included.

[c] "IT Use" is operationalized as the number of times a student logged into BIStro each week.

[d] Female = 0, Male = 1.

[e] Ages of participants ranged from 19 to 46.

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$

Data Testing

Our assumption tests provided valuable information for developing our growth models. For example, the boxplots and scattergrams showed that BIStro usage dropped to an unusually low level during Week 10—the week of spring break—so this week was considered an outlier and dropped from the study. In addition, network problems in Week 7 led us to drop data from that week as well. While these deletions improved the quality of the data set with little chance of altering the results of the study, they precluded the use of LGM because the resulting data set was not regularly sequenced (i.e., it was not properly "time-structured").

Boxplots of usage behaviors over time also showed that the participants' BIStro use over the semester exhibited nonlinear qualities. Closer examination showed that the nonlinearity corresponded to two distinct linear patterns. The first pattern involved a linear rise in the average number of log-ins to BIStro over Weeks 1–8. The second pattern involved a sudden drop in BIStro use at Week 9, followed by a linear rise in use to the end of the semester. This discontinuity in BIStro use between Weeks 8 and 9 coincided with the timing of mid-term exams, a pattern of behavior described in previous research [Brotherton and Abowd 2004]. Following recommendations in Singer and Willet [2003: 206, 233], we divided the data set into two "epochs" corresponding to these two linear patterns. This separation also permitted an assessment of potential feedback effects from the midterm exam. These patterns offered preliminary support for Hypothesis 1.

The boxplots also showed the *BIStroUse* data was positively skewed, which is typical of count data such as ours. As is commonly done in other regression methods, we mitigated the skewness by creating a new variable, *Log10Use*, which contained the logarithmic transformation *BIStroUse = log(1+BIStroUSe)*. The effects of this transformation are displayed in Figure 2, which is based on data from Epoch 2 (Weeks 9–17).

The division of our data into two epochs created a potential problem for estimating the intercept of the second epoch. If the Week variable were retained, then the intercept for Epoch 2 would be far outside that epoch's data range of $9 \le \text{Week} \le 17$. We solved this problem by creating a new variable, Time, that represented the passage of time throughout each data set. We converted week numbers for data captured in the first half of the semester (i.e., Epoch 1, Weeks 1, 2, …, 6, 8) to Time = 0, 1, …, 5, 7. We converted week numbers for data captured in the second half of the semester (i.e., Epoch 2, Weeks 9, 11, 12, …, 17) to Time = 0, 2, 3,…, 8. While this decision provided meaningful intercepts for each epoch under the theoretical and experimental contexts of our study, it also required that we analyze each epoch separately to avoid confounding them.

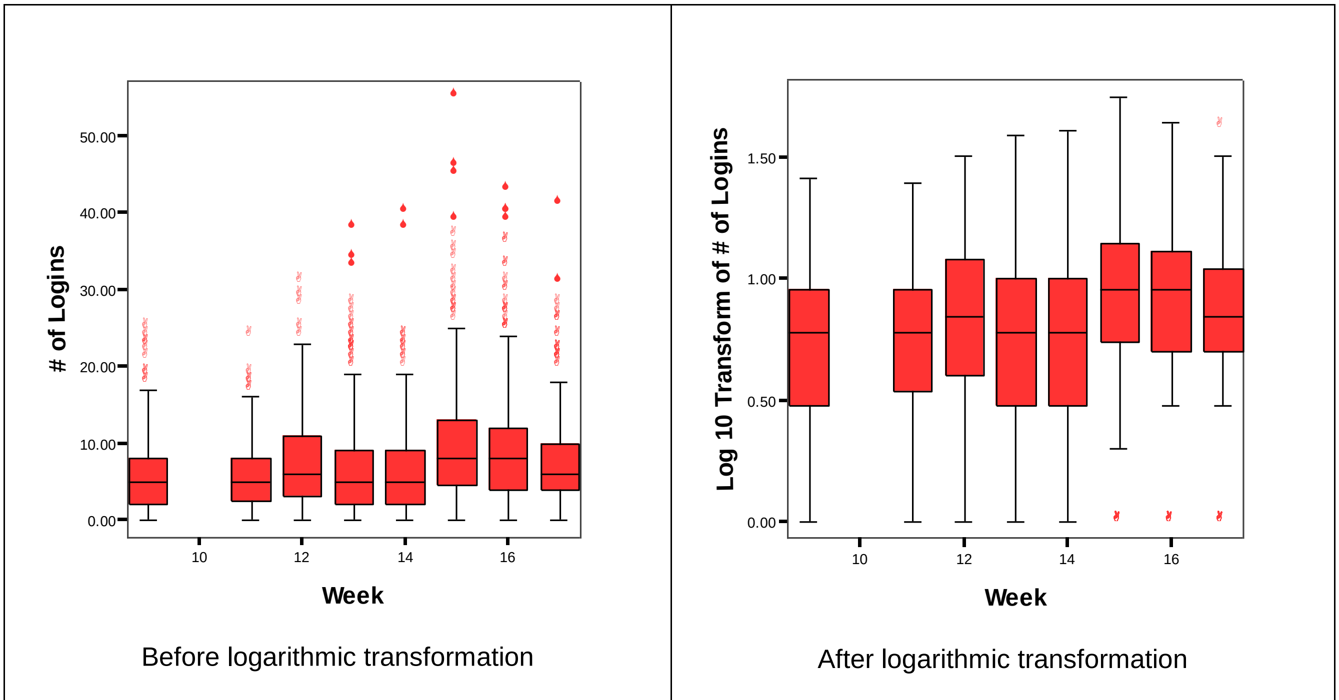


Figure 2. Effect of Logarithmic Transformation on Dependent Variable (Weeks 9–17)

Preliminary ANOVA and Regression Tests

Preliminary analyses using traditional ANOVA and OLS regression are useful because they provide insights into the level-1 equation (i.e., Equation 3). Results from the ANOVA and OLS regression analyses (Tables 2a–b) provide baselines from which to measure the usefulness of growth modeling. Results from One-way ANOVA tests showed significant but weak relationships between *Log10Use* and gender and age, as measured by η^2. The η^2 values describing the association between *Log10Use* and GPA were high, but that is not surprising, given that each of the roughly 169 GPA values was treated as a group in One-way ANOVA, and not as a measurement on continuous scale. OLS regression tests were then run on *Log10Use* and *Time* to explore that association further. Results from the OLS regression tests displayed in Table 2b showed significant but weak relationships between *Time* and *Log10Use* (adjusted R^2 = 0.03 for Weeks 1–8 and 9-17).

V. CONSTRUCTING THE GROWTH MODELS IN *HLM FOR WINDOWS*

Constructing growth models in *HLM for Windows* is similar in many respects to the way other random coefficient models are constructed in this software package. Tutorials for constructing “typical” random coefficients models are available elsewhere [e.g., Scientific Software International 2008; The University of Texas at Austin 2008; Raudenbush et al. 2004], so we will concentrate on the differences involved in growth modeling.

One of the first decisions in constructing a growth model reflects the character of collected data. Datasets with no missing data can use hierarchical linear model (HLM). Datasets with randomly missing data—as in studies that aimed to collect *T* observations/person, but collected only n_j observations/person ($n_j \le T$)—would be better served by using hierarchical multivariate linear model (HMLM) [Raudenbush et al. 2004: 140]. Since we had no missing data, we used the two-level hierarchical linear model (HLM2) choice in the “Select MDM type” window (Figure 3).

Table 2: Results of ANOVA and OLS Regression Analyses for BIStro Use

Weeks[a]	Dependent Variable	Factor[a]	n[b]	*F*	η^2	df
1–8	Log10Use[c]	Gender	2,429	80.175	0.032***	1
		Age	2,422	4.525	0.029***	16
		GPA	2,429	4.480	0.251***	169
9–17	Log10Use[c]	Gender	2,776	14.539	0.005***	1
		Age	2,768	7.348	0.041***	16
		GPA	2,776	3.648	0.191***	169

(a)
One-way ANOVA

Weeks[a]	Dependent Variable	Coefficients	n[b]	*Adj. R^2*	Unstandardized Coefficients B (Std. Error)	df
1–8	Log10Use[c]	Constant	2,429	0.03***	0.615 (0.014)	2,428
		Time			0.034 (0.004)	
9–17	Log10Use[c]	Constant	2,776	0.03***	0.685 (0.014)	2,775
		Time			0.024 (0.003)	

(b)
OLS Regression

[a] Week numbers were converted to *Time* values to permit meaningful intercepts. Weeks 1–8 were converted to Time 0–7; Weeks 9–17 to *Time* 0–8 as well, but analyzed separately to avoid confounding. Data from Weeks 7 and 10 were discarded due to a network problem and spring break, respectively.

[b] n is based on the number of students times the number of weeks of collected data.

[c] "Log10Use" is operationalized as the base 10 logarithm of the number of times a student logged into BIStro each week.

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

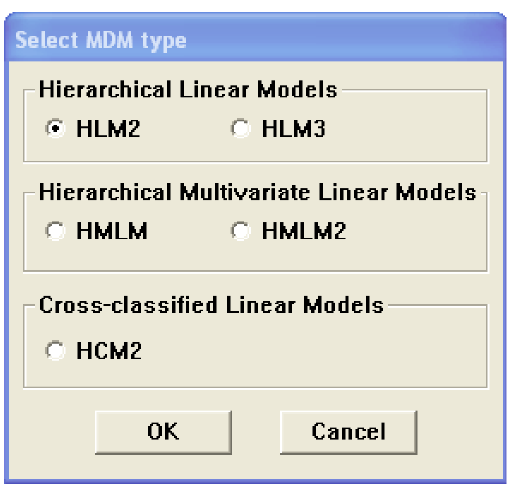


Figure 3. "Select MDM type" Window in *HLM for Windows*

Another important difference in growth modeling involves the nesting of the event-level of analysis within the person-level of analysis. Choosing the "measures within persons" selection in the "Make MDM" window (Figure 4) achieves this goal as well as ensuring the use of accepted variable terminology (i.e., π and e for the level-1 equation and β and r for the level-2 equations).

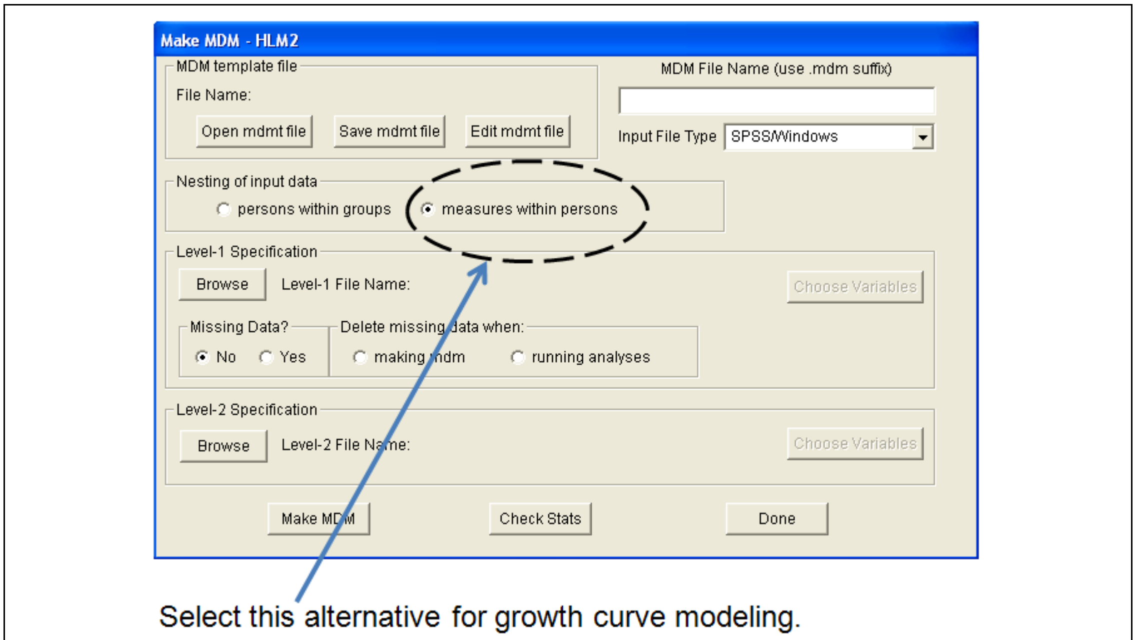


Figure 4. "Make MDM" Window in *HLM for Windows 6.06*

The Unconditional Means Model

The first model created is called the *unconditional means model* [Singer and Willett 2003], also known as the *fully unconditional model* [Raudenbush and Bryk 2002: 24] and the *empty model* [Snijders and Bosker 1999: 45]. It consists of the dependent variable Y_{ti}, the level-1 and level-2 intercepts (i.e., π_{0i} and β_{00}, respectively) and the level-1 and level-2 error terms (i.e., e_{ti} and r_{qi}, respectively). The lack of predictor variables permits the mean of the dependent variable and the level-1 and level-2 error to be calculated unconditionally. The unconditional means model thus serves as a baseline from which to measure the usefulness of subsequent models. Following Equations 3 and 4, the level-1 and level-2 models for the unconditional means model are as follows:

$$Y_{ti} = \pi_{0i} + e_{ti} \tag{5a}$$

$$\pi_{0i} = \beta_{00} + r_{0i} \tag{5b}$$

Equations 5a–b can now be used to explain the derivation of the name *unconditional means model*. This name originates from the implicit modeling of slopes in the level-1 equation (i.e., Equation 5a). The notion of "means" derives from the intercept values π_{0i} in Equation 5a, each of which represent "the true mean of Y for individual i" [Singer and Willett 2003: 92]. Since π_{0i} has no level-2 predictors, its estimation is *unconditional*.

Building the unconditional means model in *HLM for Windows* is similar to building other types of random coefficient models [e.g., Scientific Software International 2008]. After the variables have been selected and the template and MDM files saved, the *HLM for Windows* model window will appear. The two-level unconditional means model (i.e., Equations 5a, b) is now entered (Figure 5). We used full maximum likelihood estimation to measure the fit of the entire model, which in turn permits likelihood ratio tests that use chi-square comparisons of changes in deviance across nested models [Singer and Willett 2003: 116–120]. Full maximum likelihood estimation is chosen by selecting Other Settings > Estimation Settings > Full Maximum Likelihood. We also selected the homoscedasticity test (Other Settings > Hypothesis Testing > Test homogeneity of level-1 variance) and level-1 and level-2 residual files (go to Basic Model Specifications, then select the Level-1 and Level-2 Residual File command buttons) to test key assumptions for two-level HLM. The unconditional means model can be constructed by selecting the dependent or outcome variable in the left pane (i.e., *LOG10USE*). The level-1 and level-2 models will be constructed automatically. Note the variable nomenclature used in *HLM for Windows* differs somewhat from Equations 5a and 5b. There is only one subscript for the level-1 intercept π_0, and no subscripts for the level-1 or level-2 residual variables e and r, respectively.

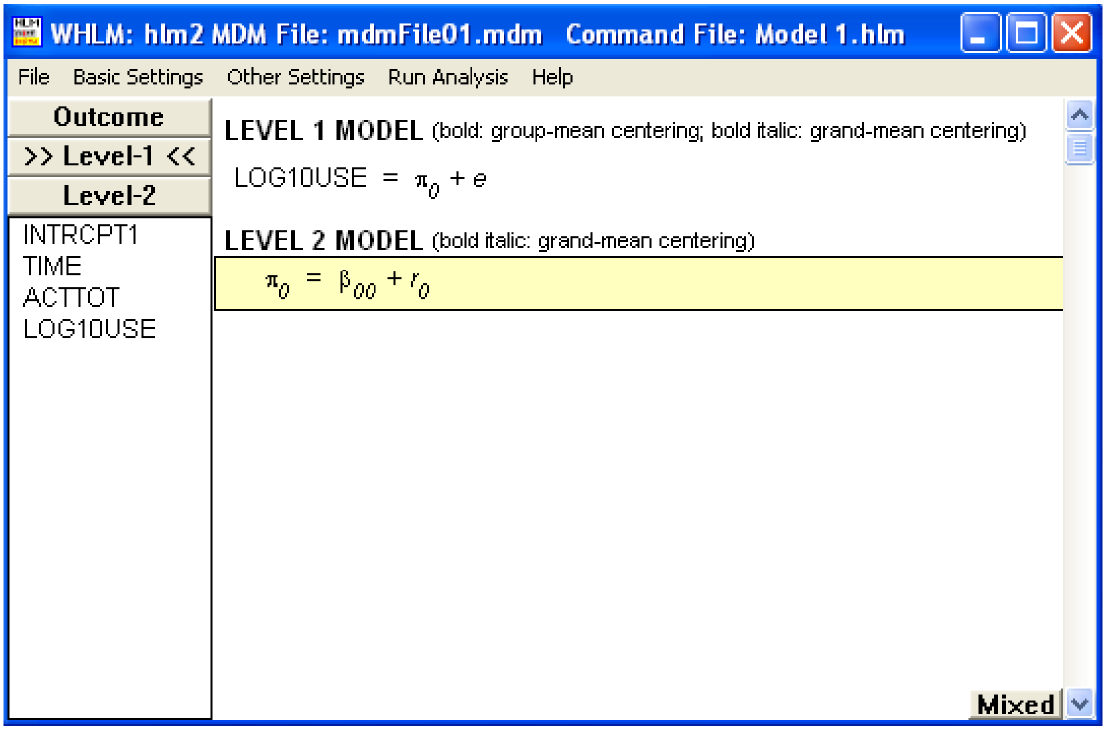


Figure 5. Unconditional Means Model (a.k.a. Fully Unconditional Model) in *HLM for Windows 6.06*

The unconditional means model can now be run and its output analyzed, just as in other random coefficients models. The run will produce an output file (hlm2.txt). We will discuss the contents of this file after showing how succeeding models are built.

The Unconditional Growth Model

In most multilevel research, the second model adds one or more independent variables to the level-1 equation (i.e., Equation 1). Growth curve models may also use multiple level-1 predictors, though it is recommended that the second model contain only one variable related to time because doing so establishes (1) if there is sufficient variation in the outcome variable "worth exploring" and (2) "*where* that variation resides (within or between people)" [Singer and Willett 2003: 92]. Since the effects of time are not conditioned upon other predictors in this simplified model, it is an *unconditional growth model* [Singer and Willett 2003].

The above approach differs from typical HLM, which may often build *random-intercept models* at this point [Snijders and Bosker 1999: 49; Raudenbush and Bryk 2002: 26]. Random-intercept models are characterized by the addition of level-2 predictors of the level-1 intercept in unconditional means models (e.g., π_{0i} in Equation 5b). We chose to model *Time* unconditionally (i.e., using an unconditional growth model) because the BIStro designers were primarily interested in the effects of gender, age, and GPA on BIStro use after controlling for time. The use of *Time* as a control variable can be seen in the wording of our hypotheses (e.g., Hypothesis 2a: *Gender directly relates to web site usage after controlling for time*).

The above approach would *not* be appropriate for researchers who are primarily interested in the effects of time after controlling for other level-2 effects (e.g., *Time directly relates to web site usage after controlling for gender*). In such cases, the random-intercept model would be constructed as the second model. Subsequent models including time as a level-1 predictor would be *conditional growth models* because time's effect would be conditioned upon level-2 predictors present in the random-intercept model.

The addition of *Time* as the only level-1 predictor in unconditional growth models provides a test of Hypothesis 1. It also creates a new coefficient, π_{1i}, which in turn requires an additional level-2 regression equation (i.e., Equation 6c). The new level-2 regression equation contains β_{10}—the mean *Time* slope across individuals—and r_{1i}, an error term associated with the *Time* slope across individuals. Our unconditional growth model takes the following form:

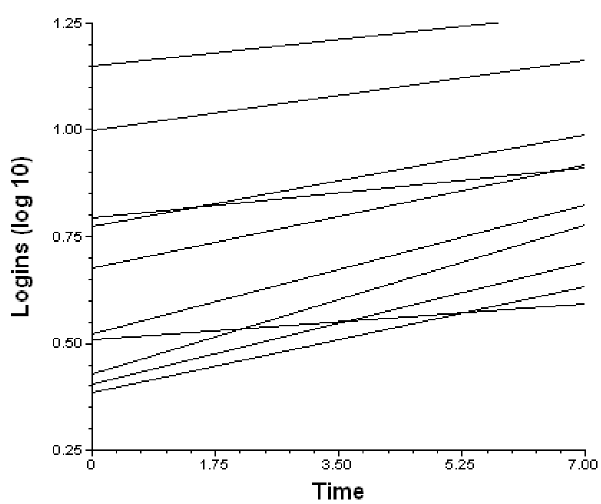
$$Log10Use_{ti} = \pi_{0i} + \pi_{1i} Time_i + e_{ti} \quad (6a)$$
$$\pi_{0i} = \beta_{00} + r_{0i} \quad (6b)$$
$$\pi_{1i} = \beta_{10} + r_{1i} \quad (6c)$$

Note that while π_{0i} and π_{1i} in Equations 6a-c each have an identifying subscript to distinguish their roles as the intercept and the *Time* coefficient (i.e., 0 and 1, respectively), the identifying subscript for the variable *Time* is unnecessary because *Time* is a named independent variable in Equation 6a.

It is important to note at this point that the researcher can model the level-1 coefficient of *Time*, β_{1i}, using "fixed" and/or "random" level-2 parameters or "effects" in the unconditional growth model. *Fixed effects* are parameters that "capture the systematic interindividual differences in change trajectory according to values of the level-2 [person-level] parameter(s)" [Singer and Willett 2003: 60]. The fixed effects in Equations 6b–c are the parameters β_{00} and β_{01}. *Random effects*, also called *stochastic components* [Singer and Willett 2003: 61], are the residuals that "represent those portions of the level-2 outcomes—the individual growth parameters—that remain 'unexplained' by the level-2 predictor(s)" [Singer and Willett 2003: 61]. The random effects in Equations 6b–c are the residuals r_{0i} and r_{1i}.

In addition to naming two types of effects, the terms *fixed* and *random* also describe assumptions the researcher must make about the variance of those effects across higher level units. Effects that are assumed to vary randomly will be predicted by higher level equations containing a residual or "random effect" [Raudenbush et al. 2008]. Thus, both so-called "fixed effects" in Equation 6a (i.e., the coefficients π_{0i} and π_{1i}) are assumed to vary "randomly" because their corresponding level-2 equations contain residuals or "random effects" (i.e., r_{0i} in Equation 6b and r_{1i} in Equation 6c, respectively). On the other hand, "fixed effects" that are assumed to be fixed or "nonrandom" are predicted by higher level equations without random effects such as r_{0i} or r_{1i}.

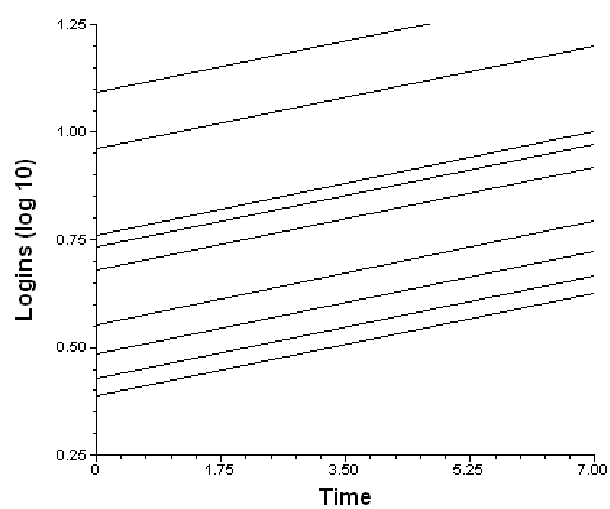
The implications of these assumptions about a coefficient's variance across higher level units are depicted in Figures 6a–b, which is based on the regression lines for the first ten participants in our sample. Note that slopes in Figure 6a—which are assumed to vary randomly and whose level-2 equations include the residual r_{1i}—show variation in slopes. The slopes in Figure 6b—which are assumed fixed and whose level-2 equations exclude r_{1i}—do not. Also notice that the intercepts change as well because the corresponding level-2 equations contain the residual r_{0i}. The fixed coefficient model on the right is not unlike the typical OLS regression, in which variation across groups is assumed to be zero. Excluding both r_{0i} and r_{1i} in *HLM for Windows* can be used to generate least squares estimates of the fixed effects β_{0i} and β_{1i}.



(a)

Slopes Assumed to Vary Randomly

(r_1 is included in higher level equation)



(b)

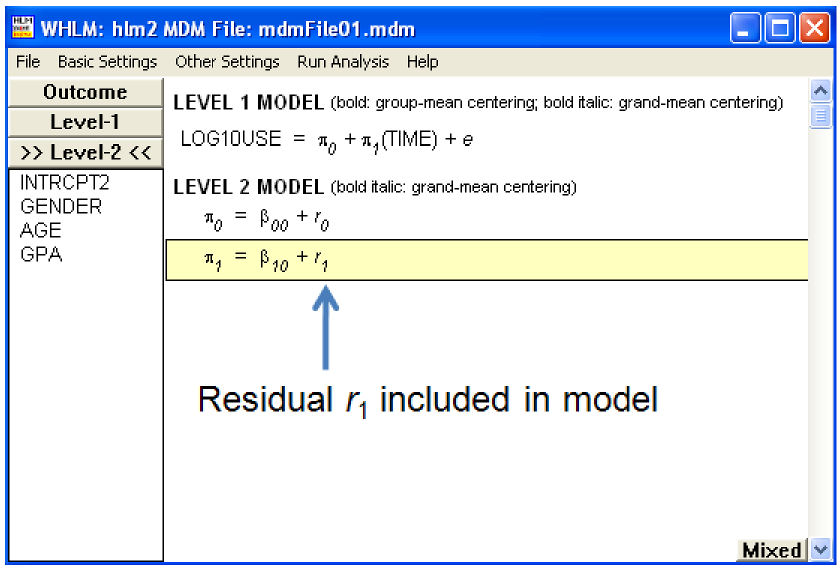
Slopes Assumed Fixed

(r_1 is excluded from higher level equation)

Figure 6. Random and Fixed Unconditional Growth Models in *HLM for Windows 6.06*

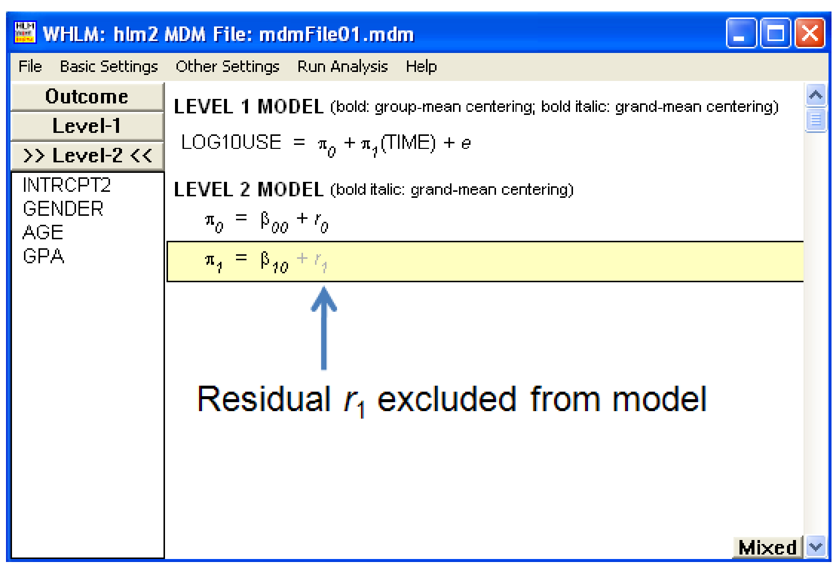
A close inspection of Figures 6a–b also shows how event-level data are nested with person-level regression lines. Each person's event-level data (i.e., *Time* and *Log10Use*) are used to construct (i.e., are nested within) a unique regression line for that person. Those regression lines vary in intercepts and slopes, thus providing two variables that describe each person's unique pattern of BIStro use. Moreover, those person-level intercepts and slopes are built upon a longitudinal series of event-level data.

The decision whether to use random *versus* fixed coefficients modeling is based on a number of factors, including "the focus of the statistical inference, the nature of the set of *N* groups, the magnitudes of the group sample sizes n_j, and the population distributions involved" [Snijders and Bosker 1999: 43]. The choice also depends on the researcher's goals and interests. If the researcher is interested in level-2 effects (e.g., the influence of individual-level characteristics in longitudinal studies), then random models will provide the requisite analysis. On the other hand, if level-2 effects are not the focus of the study, or if their addition would unnecessarily complicate the model, then a fixed model would be the simpler, more parsimonious choice. The implementation of this choice in *HLM for Windows* is depicted in Figure 7.



(a)

Random Coefficient Model



(b)

Fixed Coefficient Model

Figure 7. Screenshots of Random and Fixed Unconditional Growth Models in *HLM for Windows 6.06*

Again, the research questions and underlying theories should guide the choice of random *versus* fixed coefficient models. Since our population was drawn from a real population of users, and we wished to draw conclusions about that population and test effects of person-level (level-2) variables, random coefficient modeling was deemed appropriate.

One easy way to make this model in *HLM for Windows* is to save the unconditional means model (Model 1) as "Model 2," and then make the appropriate modifications to Model 2. This trick ensures that previous choices regarding full likelihood estimation, homogeneity of level-1 variance, creation of level-1 and level-2 residual files, etc., in the unconditional means model are not inadvertently overlooked.

Another critical decision in creating the unconditional growth model involves the centering of the *Time* variable. As in regression, centering an independent variable involves subtracting a constant—usually the variable's mean—in order to make the intercept meaningful [Cohen et al 2003: 262]. In HLM, independent variables can be grand mean or group mean centered (i.e., the subtrahend can be the mean of all cases, or just the cases of the relevant group). Grand mean centering is often recommended in typical HLM analyses because it provides computational advantages and can reduce intercept and slope estimate correlations [Hofmann and Gavin 1998]. Group mean centering removes level-2 differences, so it is not appropriate in our case.

Centering is a complex issue which cannot be thoroughly treated in this tutorial. Readers are advised to familiarize themselves with the nuances of centering. A basic discussion of centering in multiple regression can be found in Cohen et al. [2003: 261–267]. Hofmann and Gavin [1998] is an oft-cited work (over 150 times in Web of Science) that contains a simulation of the different effects of grand mean *versus* group mean centering. Enders and Tofighi [2007] is another excellent resource, with a helpful discussion on the linkage between centering and research questions (pages 127–134).

We felt that raw matrix centering was preferable in our particular case because we were interested in how students used BIStro at the beginning of each epoch (i.e., at the beginning of the semester and in the first week after the midterm exam). We were able to enter *Time* "uncentered" (i.e., using raw matrix centering) because the value "0" was made meaningful in both epochs when the *Week* data, ranging from 1–8 and 9–17 in Epochs 1 and 2, respectively, was linearly transformed to the *Time* variable (i.e., to 0–7 and 0–8 in Epochs 1 and 2, respectively). The entry of *Time* as an uncentered variable is signified in *HLM for Windows* by normal font.

The decision about whether level-1 coefficients are fixed or vary randomly (Figure 6) can now be implemented. Coefficients that are assumed to vary randomly in *HLM for Windows* are made by toggling the corresponding residual (e.g., r_1) to the "on" position, which is signified by the normal (i.e., darker) grayscale font color (Figure 7a). Coefficients that are assumed fixed are created by toggling the corresponding residual to the "off" position, which is signified by a lighter grayscale color (Figure 7b). It is important to check these residuals before runtime because researchers may inadvertently toggle the residual to the wrong position as they work with their models over time. Researchers and reviewers can check the correct setting by observing changes in parameters across subsequent models, which will be discussed shortly in more detail in the *Goodness of Fit* subsection.

The unconditional growth model can now be run. Detailed explanations of the *HLM for Windows* output can be found elsewhere, including textbooks on HLM [e.g., Raudenbush and Bryk 2002] and web sites of university statistics help desks [e.g., The University of Texas at Austin 2008]. Outputs for the test of homogeneity of level-1 variance, level-1 residuals, and level-2 residuals—selected earlier during the building of the unconditional means model—should also be examined. Output for the test of level-1 variance homogeneity is typically found at the end of the *HLM for Windows* output file. The results from our test of the unconditional means model (Figure 8) show that the chi-square statistic is not significant, indicating the null hypothesis of level-1 variance homogeneity is accepted.

```
Test of homogeneity of level-1 variance
----------------------------------------
Chi-square statistic         =     47.45204
Number of degrees of freedom =  342
P-value                      = >.500
```

Figure 8. Results from Homoscedasticity Test

We also examined the level-1 and level-2 residual files (i.e., resfil1.sav and resfil2.sav, respectively) using *SPSS 15.0 for Windows*. This examination included the construction of Normal Q-Q plots of the level-1 residuals—created by selecting Analyze > Descriptive Statistics > Q-Q Plots—which supported the assumption of residual normality. Other tests are explained in *HLM for Windows'* Help section under "Model checking based on the residual file." We will not view the output of the hlm2.txt file at this time, but will do so later in the *Results* and *Discussion* sections when the results of all models will be compared.

Conditional Models

The next decision involves the addition of other predictors to the level-1 model. These can include polynomial functions of time [Singer and Willett 2003: 214]) or other event-related predictors. Since our study is interested only in the effects of time as a level-1 predictor, we will now move on to the construction of our two level-2 conditional models, also known as fully multivariate models [Snijders and Bosker 1999].

Our third model, represented in Equations 7a–c, allows for tests of direct effects of person-level characteristics on the event-level outcome variable Y_{ti}—via the level-1 intercept π_{00}—after controlling for time. We accomplished this goal by taking the level-2 equation associated with the level-1 intercept (i.e., Equation 6b) and adding the desired independent variables for gender, age, and academic performance, the last of which is operationalized by the student's grade point average (GPA) at the beginning of the semester. The result is Equation 7b, in which each independent variable represents a particular characteristic of a given student (e.g., $Gender_i$ represents the gender of student *i*). This structure permits the testing of Hypotheses 2a, 3a, and 4a. The level-1 equation does not change (i.e., Equation 6a = Equation 7a). The third model for the BIStro study is as follows:

$$Log10Use_{ti} = \pi_{0i} + \pi_{1i}Time_i + e_{ti} \tag{7a}$$

$$\pi_{0i} = \beta_{00} + \beta_{01}Gender_i + \beta_{02}Age_i + \beta_{03}GPA_i + r_{0i} \tag{7b}$$

$$\pi_{1i} = \beta_{10} + r_{1i} \tag{7c}$$

A comparison of Equations 6a-c and 7a-c shows why the growth curves have been built in this particular sequence. Equations 6a–c give the unconditional effect of *Time* on *Log10Use*, while Equations 7a–c give the effect of *Gender*, *Age*, and *GPA* on *Log10Use* conditioned upon *Time*. These consecutive models are consistent with the BIStro designers' interest. Moreover, because the BIStro designers were not interested in predicting BIStro use on any particular week and were satisfied with an average rate of change in BIStro use across each epoch, other more complex models offered no meaningful advantages. The model presented in Equations 7a–c is preferred given its parsimonious structure. Other researchers with other agendas may need to build their models differently. For example, researchers who are interested in predicting or discovering IT use at any given point in time would probably prefer to use fixed- or variable-occasion designs that use dummy variables to represent time [Snijders and Bosker 1999: 167–198]. Autoregressive models, which use a variable's prior values to determine a current value, may also be useful [Bollen and Curran 2004]. Again, it is the underlying theory and research questions that drive model building.

We can now explore the difference between random and fixed assumptions in Model 3 by focusing on the effects of level-2 predictors on the level intercept, π_{0i} (Figure 9). We will simplify this example by using *Gender* as the only level-2 predictor and diagram the differences between random and fixed effects (i.e., whether r_1 is included or excluded, respectively, in the HLM modeling of Equation 7c). Note the similarities in slopes and intercepts between Figures 6a and 9a and between Figures 6b and 9b. Though close, there are slight differences in intercept values, as would be expected with the addition of gender as a level-2 predictor. Because gender has been used only to differentiate intercepts, the slopes in Figure 9b are all equal (i.e., all slope coefficients have been fixed at the same value). As in Figures 6a–b, the graphs in Figures 9a–b are based on data from the first ten participants in our data set, and are not necessarily representative of the entire data set. Finally, perhaps the most noticeable difference between Figures 6a–b and 9a–b is that we chose to identify regression lines in the latter by gender. We did not use this option in Figure 6 because we wanted to simplify that presentation.

Construction of Model 3 starts by saving Model 2 as Model 3, again to ensure that our choices regarding full maximum likelihood estimation, output files, etc., are not overlooked. Equations 7a–c are entered into *HLM for Windows* as follows (Figure 10). Age and GPA are entered as grand mean centered for several reasons. First, the value "0" is not meaningful or practical in either variable, which precludes raw matrix centering. Second, the use of group mean centering would cause the variance of the level-1 intercept π_{0i} to represent between group variance only, and would not partial out the effects of added level-1 variables [Hofmann and Gavin 1998]. Third, grand mean centering of Age and GPA allows their intercepts to be meaningfully interpreted as "the expected level of the outcome for a person with an 'average' level on the predictor" [Hofmann 1997: 738], and also controls for level-1 variance in assessing level-2 variables [Hofmann 2006]. Entering Age and GPA as grand mean centered variables in *HLM for Windows* is signified by boldface italic type. Gender, on the other hand, is a dummy (0,1) variable and must therefore be entered uncentered because grand mean centering—which would use the "grand mean" of the dummy variables (i.e., about 0.5, depending on the female/male ratio)—would make the results difficult to interpret. Adding *Gender* as an uncentered variable in HLM is signified with regular type.

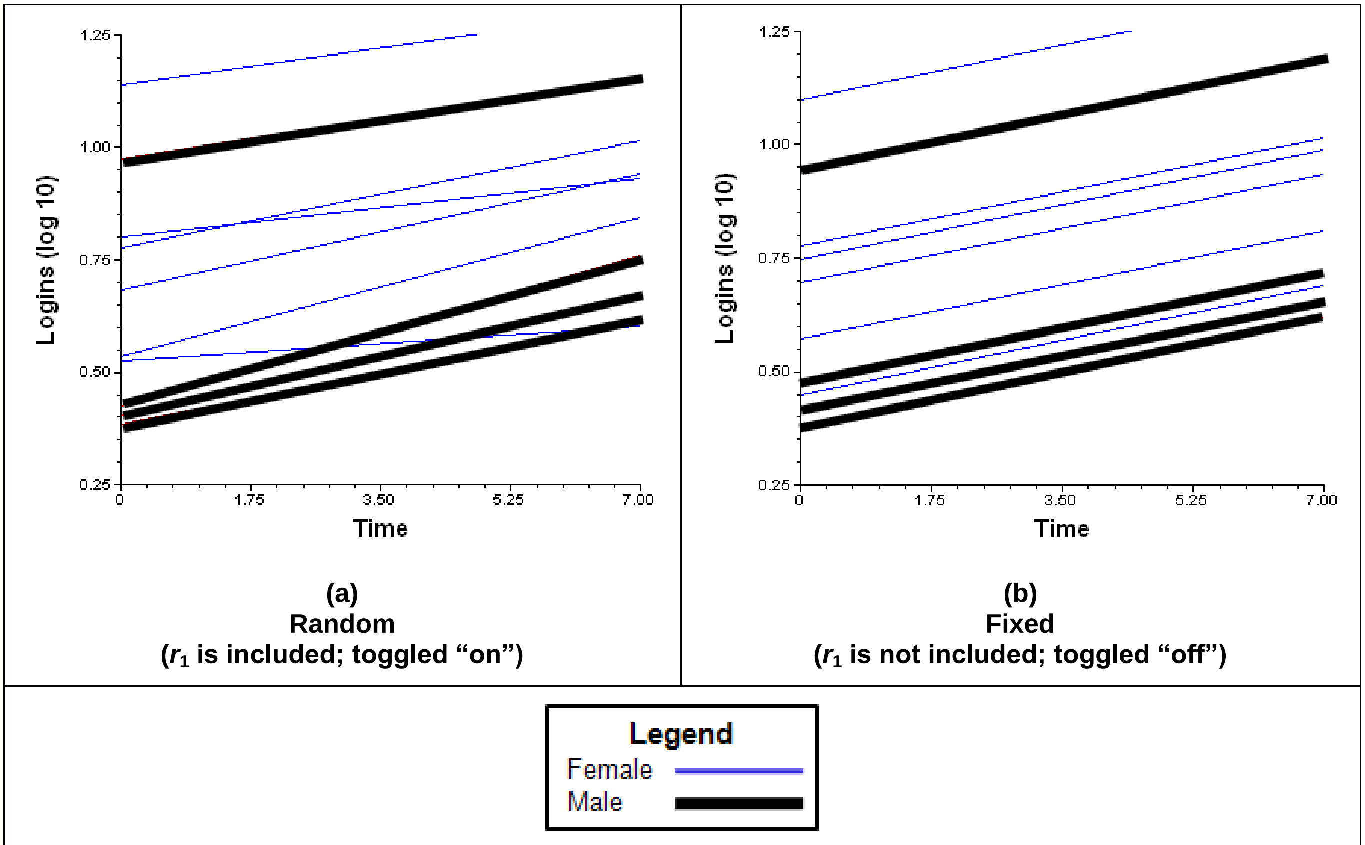


Figure 9. Random and Fixed Effects on Level-1 Intercept, π_{0i} (Gender only, Epoch 1)

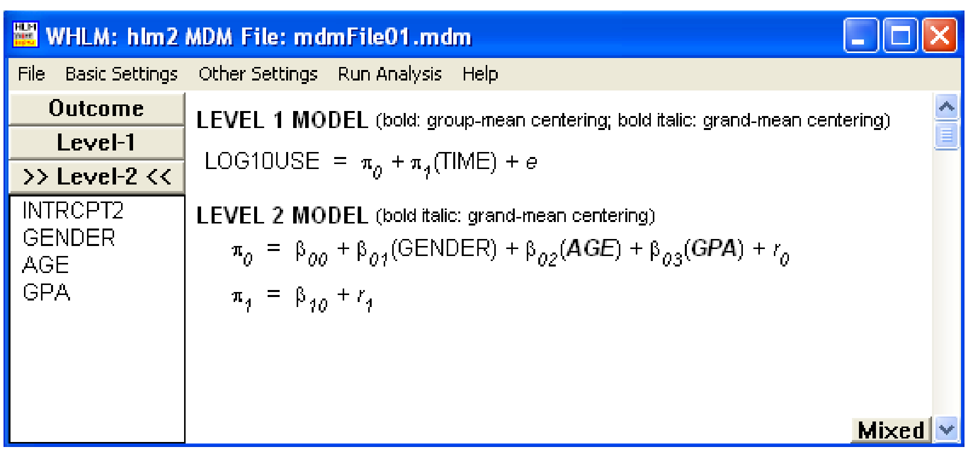


Figure 10. Screenshot of Model 3 in *HLM for Windows 6.06*

The construction of the fourth model (Figure 11) is similar to that of the third. Moderation effects on the relationships between *Time* and *Log10Use* are modeled by taking the level-2 equation associated with the coefficient of *Time* (i.e., Equation 6c or 7c—they're both the same) and adding desired predictors to produce Equation 8c. No changes are necessary for Equations 7a or 7b (i.e., Equation 7a = Equation 8a; Equation 7b = Equation 8b). Equation 8c now represents the rate of change effects of gender, age, and GPA on the relationship between *Log10Use* and *Time* (i.e., on the slope coefficient π_{1i}). The structure facilitates the testing of Hypotheses 2b, 3b, and 4b. As in Model 3, Gender was added uncentered, as signified by the variable's "normal" font. Age and GPA, on the other hand, were entered using grand mean centering, as signified by the boldface italic font. Again, identifying subscripts are dropped for named predictor variables, but retained for their corresponding coefficients.

$$Log10Use_{ti} = \pi_{0i} + \pi_{1i}Time_i + e_{ti} \quad (8a)$$
$$\pi_{0i} = \beta_{00} + \beta_{01}Gender_i + \beta_{02}\boldsymbol{Age_i} + \beta_{03}\boldsymbol{GPA_i} + r_{0i} \quad (8b)$$
$$\pi_{1i} = \beta_{10} + \beta_{11}Gender_i + \beta_{12}\boldsymbol{Age_i} + \beta_{13}\boldsymbol{GPA_i} + r_{1i} \quad (8c)$$

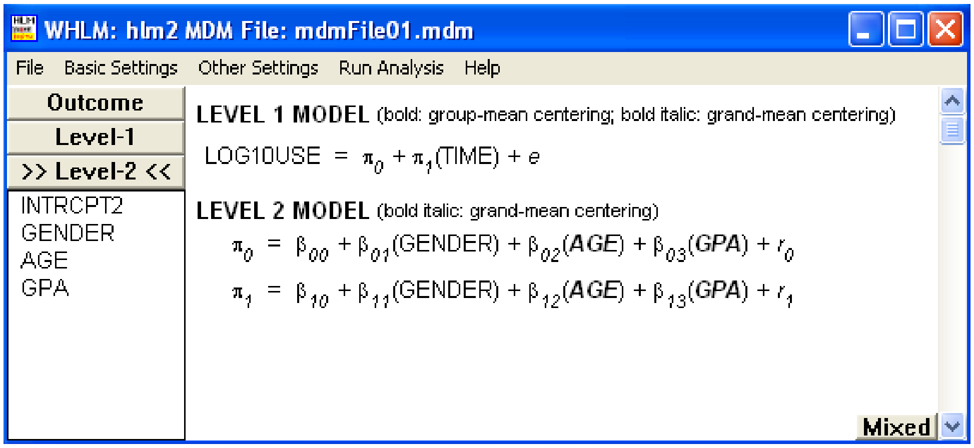


Figure 11. Screenshot of Model 4 in *HLM for Windows 6.06*

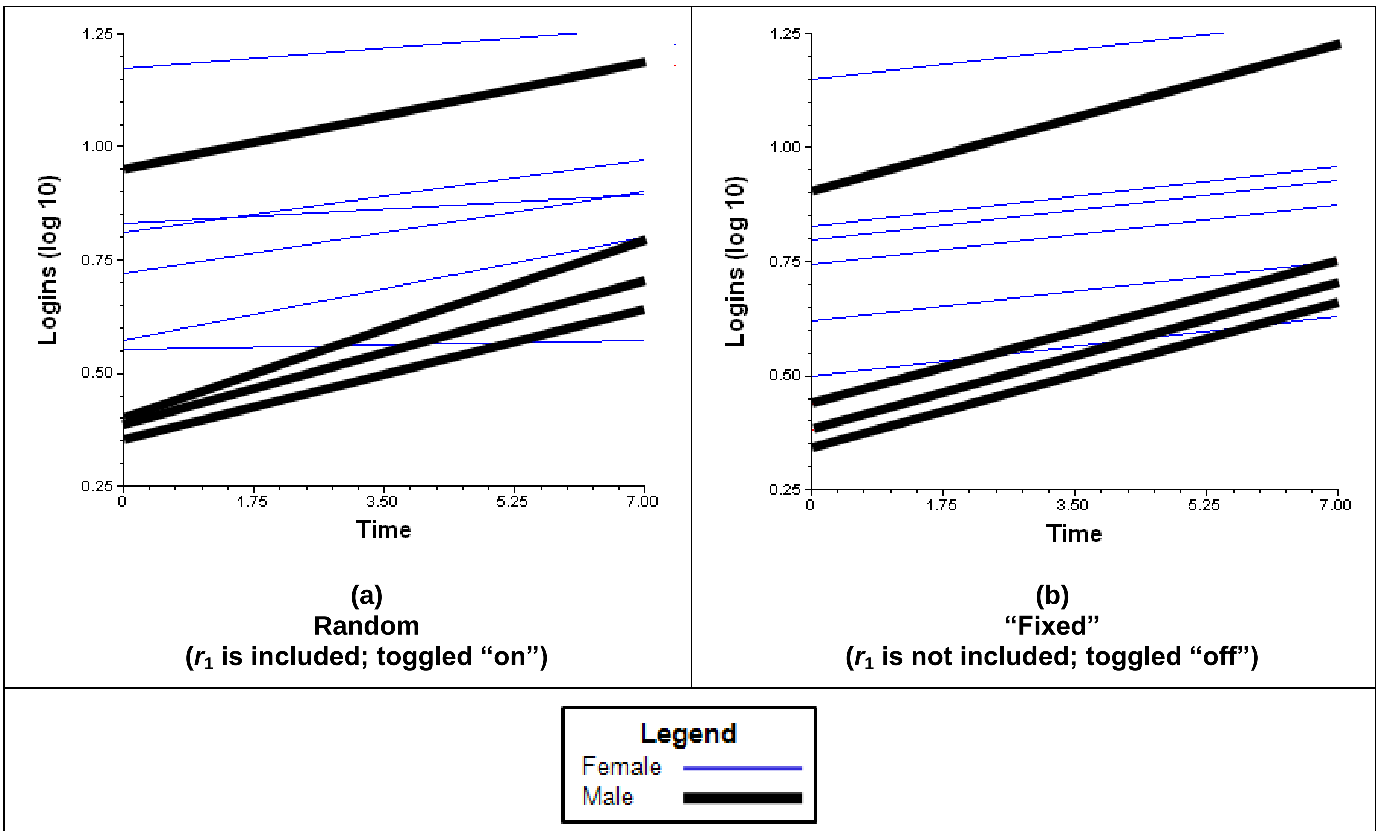


Figure 12. Random and Fixed Assumptions on Level-1 Intercept (π_0) and Time Slope (π_1) (Gender only, Epoch 1)

As before, the level-1 coefficient of Time, π_{1i}, can be assumed to be fixed or vary randomly across subjects. The difference between these assumptions is depicted in Figures 12a–b, which builds upon Figures 9a–b. As in Figure 9, Figure 12 shows the regression lines for the first ten participants in our sample. Figures 12a–b demonstrate the effects of the level-2 predictor GENDER on π_0 and π_1 (i.e., Age and GPA not included in this example).

A comparison of Figures 9a–b with 12a–b shows differences in intercepts and slopes, as would be expected with the addition of Gender as a level-2 predictor of π_1. For example, in Figures 9a and 12a, there is a flattening in the slope of the highest female regression line and an increase in the slopes of the bottom three male regression lines. A comparison of Figure 9b and 12b also shows the "fixed" assumption in the latter is now conditional on gender; that is, all lines in Figure 9b have the same slope, while the lines in Figure 12b are divided into two sets (i.e., male, represented by the darker lines, and female, represented by the lighter lines), each set with its own unique slope. Again, researchers should also check that they have not toggled the residuals accidentally before running their models.

Goodness of Fit

The goodness-of-fit tests (i.e., likelihood ratio tests) measure changes in deviance (–2 log likelihood), which shows "*how much worse* the current model is in comparison to the best possible model" and "is identical to the residual sum of squares" in regression analysis [Singer and Willett 2003: 116]. Likelihood ratio tests are well described elsewhere [e.g., Singer and Willett 2003: 116–122], so we will not belabor the point here. However, a few relevant points are worth mentioning.

First, deviance measures of two models can be compared only when both models are based on the same data set and when one of the models is nested within the other. A second point focuses on ensuring the likelihood ratio tests are based on the models actually described in a manuscript. One way to accomplish this is by checking if the processions of parameters reported in *HLM for Windows* results (e.g., the reported parameters 3, 6, 9, and 12 in Table 3) are consistent with reported estimation procedures and fixed *versus* random assumptions about level-2 effects. Researchers should check the *HLM for Windows* output file to ensure the correct estimation method and model equations have been specified. The estimation method is especially easy to miss because full maximum likelihood estimation, which is used in many model comparisons, is not the default choice in *HLM for Windows* or many other RCM software packages.

Reviewers, who often do not see these output files, can rely upon the procession of parameter counts to see if the reported estimation methods and model equations are consistent with reported results. The number of parameters used in a model can be calculated by adding the number of fixed effects and the number of variance–covariance components, the latter of which is equal to "$m(m + 1)/2 + 1$, where m equals the number of random effects in the level-2 model" [Raudenbush and Bryk 2002: 84]. Thus, our Model 1D has eight fixed effects (i.e., β_{00}, β_{10}, β_{01}, β_{02}, β_{03}, β_{11}, β_{12}, and β_{13}) and four variance-covariance components (i.e., r_{0i} and r_{1i}; $\{2 \times (2 + 1) / 2\} + 1 = 4$) for a total of twelve parameters. Changing either the estimation method or random *versus* fixed assumption can therefore change the number of parameters. We can use Model 1D as an example. If we had used restricted maximum likelihood estimation in Model 1D, the parameter count would have changed from 12 to 4 because the effects of the eight fixed variables (i.e., β_{00}, β_{10}, β_{01}, β_{02}, β_{03}, β_{11}, β_{12}, and β_{13}) would have been removed. If we kept full maximum likelihood estimation in Model 1D but had instead assumed fixed variation in our modeling of π_1, (i.e., removed r_{1i}), the number of parameters would have been reduced by *two* because there would be only one residual left (i.e., r_{0i}; $\{1 \times (1 + 1) / 2\} + 1 = 2$). It is therefore critical that researchers report the number of estimated parameters used in each growth model.

Variance and Variance Explained Statistics

Results from multilevel modeling software such as *HLM for Windows* include measures of level-1 variability, σ^2, and level-2 variance-covariance measures (e.g., τ_{00}, τ_{01}, and τ_{11}). The statistics τ_{00} and τ_{11} measure the variance of level-1 intercepts and slopes, respectively, while τ_{01} measures their covariance. Note the absence of a superscript "2" in τ_{00}, τ_{01}, and τ_{11}. These statistics can then be used to calculate a number of measures of variance explained.

The *intraclass correlation coefficient*, or ρ, indicates the theoretical maximum proportion of total variance attributable to variance between level-2 entities. The ρ statistic is calculated by dividing total variance (i.e., the sum of level-1 and level-2 variances) by the amount of level-2 variance (Equation 9). The statistics $\sigma_{00}{}^2$ and τ_{00} in Equation 9 represent within- and between-person residual variance, respectively, in the unconditional means model.

$$\rho = \frac{\tau_{00}}{\sigma^2 + \tau_{00}} \qquad (9)$$

Table 3: Results of HLM Estimations for *Log10 Use* over Time [a]

Variables (Coefficients)	Weeks 1–8				Weeks 9–17			
	Model 1A	Model 1B	Model 1C	Model 1D	Model 2A	Model 2B	Model 2C	Model 2D
Level-1 Intercept and *Time* Variable								
Intercept (π_{00})	0.722*** (.015)	0.616*** (.020)	0.693*** (.025)	0.748*** (.027)	0.790*** (.012)	0.685*** (.016)	0.709*** (.021)	0.760*** (.023)
Time (π_{10})		0.034*** (.003)	0.034*** (.003)	0.018*** (.005)		0.024*** (.002)	0.024*** (.002)	0.013*** (.003)
Level-2 Effects of Gender, Age, and GPA								
Direct Cross-Level Effects of Gender, Age, and GPA on Level-1 Intercept π_{00}								
Gender [b] (β_{01})			-0.132*** (.028)	-0.227*** (.037)			-0.040† (.024)	-0.128*** (.031)
Age [c] (β_{02})			0.016* (.006)	0.024*** (.007)			0.013* (.005)	0.017** (.006)
GPA [c] (β_{03})			0.043† (.026)	0.031 (.034)			0.048* (.021)	0.073** (.027)
Moderating (Rate of Change) Cross-Level Effects of Gender, Age, and GPA on Level-1 Time Slope π_{01}								
Gender [b] (β_{11})				0.027*** (.007)				0.019*** (.003)
Age [c] (β_{12})				–0.002* (.001)				–0.001 (.001)
GPA [c] (β_{13})				0.004 (.006)				–0.005 (.004)
Pseudo-R^2 Statistics[d] and Goodness-of-fit								
ρ	0.33				0.30			
R_e^2		0.11	0.11	0.11		0.05	0.05	0.05
R_r^2			0.17	0.20			0.12	0.12
Deviance	2,205.21	2,061.26	2,028.52	2,009.21	1,764.53	1,651.73	1,637.03	1,615.41
Δ Deviance		–143.95***	-32.74***	-19.31**		–112.80***	–14.70**	–21.62**
Parameters	3	6	9	12	3	6	9	12

[a] The level-1 dependent variable is the $\log_{10}$ of the number of BIStro logins per week. For Weeks 1-8, the level-1 N = 2,422 and the level-2 N = 346. For Weeks 9-17, the level-1 N = 2,768 and the level-2 N = 346. Unstandardized coefficient estimates and robust standard errors (in parentheses) reported.

[b] Gender: Female = 0, Male = 1. Entered uncentered.

[c] Entered grand mean centered.

[d] Pseudo-R^2 and ρ statistics are described in Equations 10 and 11. ρ is the intraclass correlation coefficient. R_e^2 and R_r^2 are the level-1 and level-2 variance explained statistics, respectively. Deviance is the -2 log likelihood.

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ † $p < 0.10$

The ρ statistic also gives an average measure of residual autocorrelation between any pair of composite residuals [Singer and Willett 2003: 97]. This can be seen by noting that each left subscript for π, β, and r in the level-2 equation $\pi_{0i} = \beta_{00} + r_{0i}$ from the unconditional means model (i.e., Equation 3b) equals zero. The zero subscript indicates not only that the equation describes the level-1 intercept, but also that the equation contains no other predictors of *event-level* variation. Since r_{0j} (an error variable) is the only variable accounting for variation *across* events, it thus describes the amount of error autocorrelation. Autocorrelated residuals are often caused by omitted predictors [Singer and Willett 2003: 85]. In addition, auto-correlated residuals do not influence fixed effects, though they do influence the precision of standard errors [Singer and Willett 2003: 264].

The value $(1 - \rho)$ does *not* necessarily give the amount of level-1 variance. This is because higher-level variables (e.g., class sections and teachers in the BIStro study) may account for some of that variance. Three- or four-level hierarchical models would be required to test for those possibilities. Such modeling is difficult and beyond the scope of the present tutorial.

A number of "pseudo-R^2" statistics have been proposed as measures of explained variance. In growth models, these include pseudo-R_e^2 and pseudo-R_r^2, which indicate the amount of variance attributable to changes in within- and between-person variance, respectively, across successive models [adapted from Singer and Willett 2003: 103–104]. The statistics σ_s^2 and τ_s represent within- and between-person variance in the successive model, *s*, respectively.

$$\text{Pseudo-}R_e^2 = \frac{\sigma_{00}^2 - \sigma_s^2}{\sigma_{00}^2} \tag{10}$$

$$\text{Pseudo-}R_r^2 = \frac{\tau_{00} - \tau_s}{\tau_{00}} \tag{11}$$

Unlike the variance measures in OLS regression equations, which are based on *one* variance component (i.e., R^2) and *one* error term, hierarchical linear models contain *multiple* variance components based on *multiple* error terms (e.g., e_{ti}, r_{0i}, and r_{1i} in Equations 8a–c, respectively). Comparing and interpreting these multiple measures led to many disagreements among statisticians about how to use these multiple measures of variance to construct meaningful measures of variance explained. A common problem with pseudo-R^2 statistics is that their components, σ_{00}^2 and τ_{00}, can change in meaning across successive models. In addition, pseudo-R^2 statistics typically do not consider other variance or co-variance components such as τ_{11}, or τ_{01}, which measure the predictor slope variance and the intercept slope co-variance, respectively. Researchers and reviewers must both take care when using, interpreting, or requiring pseudo-R^2 statistics [e.g., Singer and Willett 2003: 104; Snijders and Bosker 1999: 104, 123].

VI. RESULTS

Table 3 contains the results from our HLM analyses. We ran four models for each data set: Models 1A → 1B → 1C → 1D covered the first epoch (i.e., Weeks 1–8), and Models 2A → 2B → 2C →2D covered the second epoch (i.e., Weeks 9–17). The first models in each series (i.e., 1A and 2A) were the unconditional means models based on Equations 5a–b. The second models in each series (i.e., 1B and 2B) were the unconditional growth models based on Equations 6a–c. These latter models added *Time* as the level-1 independent variable, and a level-2 equation modeling π_{1i}, the *Time* coefficient, permitting a test of Hypothesis 1. The unconditional growth models are similar to—but not the same as—the OLS regression models. The difference between the two lies in the use of fixed coefficient modeling in OLS regression model *versus* random coefficient modeling in growth modeling—which requires the level-2 residuals r_{0i} and r_{1i} (i.e., Equations 6b–c, respectively). Similarities between the OLS regression and growth models can be shown in part by the similarity in B and standard error values in Table 2 against their respective counterparts for Intercept (π_{00}) and Time (π_{10}) values in Table 3.

The third models in each series (i.e., 1C and 2C), based on Equations 7a-c, accounted for direct effects of *Gender*, *Age*, and *GPA* on *Log10Use* after controlling for time. We constructed the third models by adding the *Gender*, *Age*, and *GPA* variables as level-2 predictors of the level-1 intercept, π_{0i}. The fourth models in each series (i.e., 1D and 2D), based on Equations 8a–c, accounted for the moderating (i.e., rate of change) effects of gender, age, and GPA on the relationship between BIStro use/week and *Time*. We constructed the fourth models by adding the *Gender*, *Age*, and *GPA* variables as level-2 predictors of π_{1i}, the coefficient of *Time*. The third models permitted tests of Hypotheses 2a, 3a, and 4a; the fourth models, Hypotheses 2b, 3b, and 4b.

The goodness-of-fit tests (i.e., likelihood ratio tests) show that four models exhibit improvement at the $p < 0.001$ level of significance (i.e., Models 1B, 1C, 2B, and 2C), while two exhibit improvement at the $p < 0.01$ level of significance (i.e., Models 1D and 2D). However, these goodness-of-fit results must be considered in light of changes in model parameters. For example, the GPA coefficient *decreases* in Model 1D (i.e., from 0.043, $p < 0.10$ in Model 1C to 0.031, non-significant in Model 1D), but *increases* in Model 2D (i.e., from 0.048, $p < 0.05$ in Model 2C to 0.073, $p < 0.01$ in Model 2D), Goodness-of-fit considerations can also include variance explained statistics. As expected, R_e^2, the measure of level-1 (i.e., event-level) variance explained, does not change in Models 1C, 1D, 2C, or 2D because only level-2 (i.e., individual-level) variables have been added to those models. On the other hand, R_r^2, the measure of level-2 (i.e., individual-level) variance explained does increase slightly in Model 1D, but not in Model 2D, These four statistics across models (i.e., the level-1 coefficient GPA significance, R_e^2, R_f^2, and Δ deviance) can then be interpreted by the researcher (e.g., the study was sufficiently adequate and powerful enough to detect GPA's direct effects in Epoch 2 but not in Epoch 1; however, GPA's effects in Epoch 2 were at best weak).

It must be noted that variance explained statistics in multilevel regression differ in important ways from those used in other forms of regression. For example, the addition of grand mean centered level-1 predictors in subsequent models can change the meaning of the level-1 intercept β_{0i}. In turn, this can change the meaning of τ_{00}, "the variance of the true means, [β_{0i}], about the grand mean, γ_{00}" [Raudenbush and Bryk 2002]. This change of meaning in τ_{00} is important because many—though not all—variance explained statistics incorporate τ_{00} (e.g., Equation 11 *versus* Equation 10). Consequently, changes in the meaning of τ_{00} also change the meaning of associated variance explained statistics. Raudenbush and Bryk [2002: 150] observe that researchers should therefore build their level-1 models first to avoid this problem (e.g., Models 1B and 2B). They also recommend that the same set of level-2 predictors be used for each level-2 outcome (e.g., Models 1D and 2D), with the level-1 intercept, β_{0i}, being specified first (e.g., Models 1C and 2C). Correctly interpreting the meaning of variance explained statistics is important because these statistics are often used as a measure of effect size and model specification adequacy, and thus as a justification for whether "more and better predictors should be investigated" in future research [Aguinis et al. 2009: 16].

V. DISCUSSION

Results from the OLS regression analysis in Table 2 show that BIStro use rose over time, and that time is a significant but weak predictor of BIStro use ($p < .001$, adjusted $R^2 = 0.03$ for Weeks 1–8 and 9–17). Results from the growth analyses in Table 3 are consistent with these findings, though the pseudo-R_e^2 statistics suggest that time explains a greater portion of event-level variance in Epoch 1 *versus* Epoch 2. Not surprisingly, the goodness-of-fit (χ^2) test results are significant given the large level-1 sample sizes (N = 2,422 and 2,776). The HLM and OLS regression results both support Hypothesis 1.

The ρ statistics from Models 1A and 2A show that moderate to large amounts of total variance in BIStro use/week are attributable to between-person (level-2) variance (i.e., 33 percent during Weeks 1–8, and 30 percent during Weeks 9–17). They also indicate a substantial amount of autocorrelation in our data. Whether or not to account for autocorrelation in models must be carefully considered. Since our results already showed a number of significant fixed effects for gender, age, and time, and the remaining fixed effects were close to zero, we did not account for autocorrelation.

Models 1B–D and 2B–D produced level-1 and level-2 variance-explained values that provide more detail about variation. As expected, the level-1 variance explained—as measured by pseudo-R_e^2—was consistent across Models 1B to 1D and 2B to 2D because all contain the same level-1 independent variable, *Time*. The sharp drop in level-1 variance explained from Epoch 1 to Epoch 2 (i.e., from 11 percent to 5 percent) suggests other level-1 factors come into play during the second half of the semester (and that time is not as important). Level-2 variance explained—as measured by pseudo-R_r^2—was somewhat higher in Weeks 1–8 than in Weeks 9–17 (i.e., 17–19 percent *versus* 12 percent, respectively).

Gender

We begin with Hypotheses 2a and 2b, which concern gender and BIStro use over time. We tested these hypotheses using the last three models in each data set. Model 1C shows that females were more likely to use BIStro during the first week of class than males, as evidenced by the *Gender* coefficient β_{01} on the level-1 intercept, π_{00} (i.e., β_{01} = -0.132, $p < 0.001$; female = 0, male = 1). These results support Hypothesis 2a. However, Model 1D shows that males increased their usage of BIStro at a faster rate than females over the first half of the class, as evidenced by the *Gender* coefficient β_{11} on the *Time* coefficient, π_{01} (i.e., β_{11} = 0.027, $p < 0.001$). These results, and the significant changes in deviance across models 1A–1D, support Hypothesis 2b, but contradict previous research in which females were found to be less likely to use computers than males [Fallows 2005; Malcolm et al. 2005; Papadakis 2000].

We also found unexpected results regarding the relationship between gender and BIStro use during the second half of the class. The *Gender* coefficient β_{01} in Model 2C was positive but weakly significant ($p < 0.10$) at Week 9—the intercept for the second set of data. This finding of little significant difference in BIStro use between males and females suggests that males' BIStro usage had almost but not quite "caught up" with females' usage half-way through the semester. A comparison of Models 2C and 2D produced a significantly improved model ($p < 0.01$), as measured by a χ^2 test of change in deviance. Gender's moderating effect was again significant and negative (i.e., β_{11} = –0.019, $p < 0.001$), with males' BIStro use continuing to increase.

Results from the growth modeling tests support Hypothesis 2b, but show a more complex relationship between gender and IT use than previously reported. Based on measures of *actual use over time*—not on self-reported intentions or perceptions—the results contradict previous research (i.e., that males are more likely to use computers than females) because it shows that females were more likely to use BIStro early in the semester, though males did catch up half way through the study period and continued to increase their use of BIStro.

The associations between gender and BIStro use, conditioned upon age and GPA (Table 3), are represented in part in Figures 13a–b. The two figures, though based on a limited sample (i.e., the same ten participants depicted in Figures 6, 9, and 12), nonetheless show the essence of that relationship. Males (in heavy lines) use BIStro less than females in Week 1, but come close to catching up to females midway through the semester (i.e., Weeks 7 and 8). Males continue to increase their use of BIStro in Epoch 2, catching up with females in their use of BIStro by Week 17. The use of random coefficients can be seen in the variation in slopes in each gender set, which is clearly visible in Figure 13a but less so in Figure 13b. Similar graphs can be constructed based on age and GPA, but we do not show them here, given our space limitations.

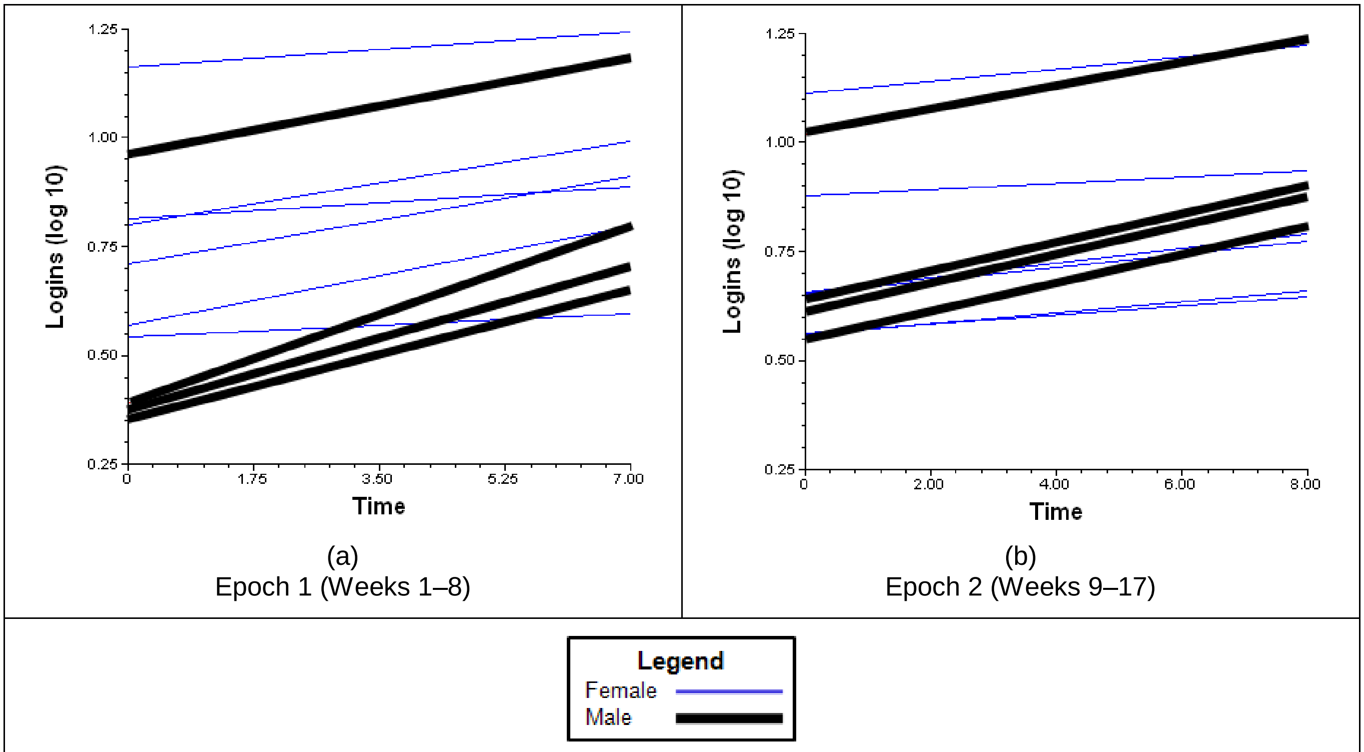


Figure 13. Random Coefficient Models over Epoch 1 and Epoch 2

The question remains, however, regarding how males and females compare in their use of BIStro at the end of the semester. Since this article is a tutorial, we took the liberty of providing a post-hoc example of how growth modeling can be used to answer this question. Following suggestions in Singer and Willett [2003: 186], we created a new level-1 variable *Backtime* in the Epoch 2 data set that would place the intercept π_{00} at the last week of that nine-week epoch. This was accomplished by a simple linear transformation of *Time*, so that the first and last weeks of the second epoch are represented by the values "–8" and "0" respectively. A new set of models were run (i.e., Models 3A–D), the results of which are displayed in Table 4. Results from Models 2A–D in Table 3 are also displayed there for comparison.

A comparison of Models 2A–D and 3A–D in Table 4 shows the effects of level-1 intercept placement. Estimations of π_{00} in Models 2A and 3A are equal, as expected, because both models rely upon β_{00}, the grand mean of Y_{ti}, as the sole predictor of π_{00}. As expected with the positive *Time* slope π_{10}, the estimations of the level-1 intercept π_{00} in Models 3B–D are higher than those of Models 2B–D. The estimations of the *Time* slope π_{10} in Models 3B–D are the same as those in 2B-D, respectively, because they are not affected by intercept placement.

Changes can also be seen in the estimations of the direct level-2 predictors gender, age, and GPA (i.e., β_{01}, β_{02}, and β_{03}, respectively). Results from Model 2D concerning the effects of gender, age, and GPA on the level-1 intercept π_{00}—positioned at Week 9, the beginning of Epoch 2—show that all three predictors have significant influence on BIStro use at the beginning of the semester ($p < 0.01$ for Age and GPA, $p < 0.001$ for Gender). On the other hand, results from Model 3D show that these same predictors have little or no direct influence on π_{00} when that intercept is positioned at Week 17 at the end of Epoch 2. Again, these changes in level-2 coefficient estimations of β_{01}, β_{02}, and β_{03} are expected because their estimations are conditioned upon estimates of the moderating effects of gender, age, and GPA (i.e., β_{11}, β_{12}, and β_{13}, respectively). The conditional effects of β_{11}, β_{12}, and β_{13} are apparent in the similar values of β_{01}, β_{02}, and β_{03} in Models 2C and 3C. It is also important to note the changes in τ, the variance in the level-1 intercepts, between Models 2B–D and Models 3B–D. Models 3B–D have less level-2 variance, which in turn affects the values of pseudo-R_r^2.

Table 4: Results of HLM Estimations for BIStro Usage over Time [a]

Variables (Coefficients)	Weeks 9–17, w/ Intercept at Week 9				Weeks 9–17, with Intercept at Week 17			
	Model 2A	Model 2B	Model 2C	Model 2D	Model 3A	Model 3B	Model 3C	Model 3D
Level-1 Intercept and *Time* Variable								
Intercept (π_{00})	0.790*** (.012)	0.685*** (.016)	0.709*** (.021)	0.760*** (.023)	0.790*** (.012)	0.877*** (.014)	0.900*** (.019)	0.863*** (.020)
Time (π_{10})		0.024*** (.002)	0.024*** (.002)	0.013*** (.003)		0.024*** (.002)	0.024*** (.002)	0.013*** (.003)
Level-2 Effects of Gender, Age, and GPA								
Direct Cross-Level Effects of Gender, Age, and GPA on Level-1 Intercept π_{00}								
Gender [b] (β_{01})			–0.040† (.024)	–0.128*** (.031)			–0.040† (.024)	0.025 (.028)
Age [c] (β_{02})			0.013* (.005)	0.017** (.006)			0.013* (.005)	0.011† (.006)
GPA [c] (β_{03})			0.048* (.021)	0.073** (.027)			0.048* (.021)	0.031 (.025)
Moderating (Rate of Change) Cross-Level Effects of Gender, Age, and GPA on Level-1 Time Slope π_{01}								
Gender [b] (β_{11})				0.019*** (.003)				0.019*** (.004)
Age [c] (β_{12})				–0.001 (.001)				–0.001 (.001)
GPA [c] (β_{13})				–0.005 (.004)				–0.005 (.004)
Variance Components								
σ^2	0.09180	0.08756	0.08764	0.08688	0.09180	0.08756	0.08764	0.08688
τ	0.04007	0.04798	0.04223	0.04199	0.04007	0.03507	0.03515	0.03544
Pseudo-R^2 Statistics[d] and Goodness-of-fit								
ρ	0.30				0.30			
R_e^2		0.05	0.05	0.05		0.05	0.05	0.05
R_r^2			0.12	0.12			0.00	–0.01
Deviance	1,764.53	1,651.73	1,637.03	1,615.41	1,764.53	1,651.73	1,637.03	1,615.41
Δ Deviance		–112.80***	–14.70**	–21.62**		–112.80***	–14.70**	–21.62**
Parameters	3	6	9	12	3	6	9	12

[a] The level-1 dependent variable is the $\log_{10}$ of the number of logins per week. For Weeks 1–8, the level-1 N = 2,422 and the level-2 N = 346. For Weeks 9–17, the level-1 N = 2,768 and the level-2 N = 346. Unstandardized coefficient estimates and robust standard errors (in parentheses) reported.

[b] Gender: Female = 0, Male = 1. Entered uncentered (i.e., raw matrix centering).

[c] Entered grand mean centered.

[d] Pseudo-R^2 and ρ statistics are described in Equations 10 and 11. ρ is the intraclass correlation coefficient. R_e^2 and R_r^2 are the level-1 and level-2 variance explained statistics, respectively. Deviance is the –2 log likelihood.

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ † $p < 0.10$

At least two conclusions can be drawn from Table 4. The first and perhaps most important is that *growth models must be carefully structured so that they answer the questions and hypotheses posed by the researcher.* The different placements of the intercepts in Epoch 2 demonstrate how this one decision can affect whether a researcher's hypotheses are accepted or rejected. Using results from Models 2C–D, one might conclude that gender is meaningfully associated with BIStro use. Using results from Models 3C–D, on the other hand, might lead one to different conclusions. It is also important for both researchers and reviewers to note the consequences of growth model structure on pseudo-R^2 statistics, and the inherent limitations of this class of statistics. A second conclusion concerns the value of growth modeling. Our study of the relationship between gender and IT use, while consistent with past research showing that "gender plays a vital role in shaping initial and sustained technology adoption decisions" [Venkatesh and Morris 2000: 129], also shows that relationship is more complex than previously thought.

Age

Hypotheses 3a and 3b concern age and IT use over time. Model 1C–D (Table 3) shows that older students were heavier users of BIStro during the first week of class than younger students (β_{02} = 0.016, p < 0.01 in Model 1C; 0.024, p < 0.001 in Model 1D). These results support Hypothesis 3a. Comparisons of Models 2B–D showed similar results. Older students were still heavier users of BIStro in Week 9 (β_{02} = 0.013, p < 0.05 in Model 2C; 0.017, p < 0.01 in Model 2D), though the difference was less significant than in Weeks 1–8. The moderating effects of age were very weak in Weeks 1–8 (β_{12} = –0.002, p < 0.05 in Model 1D) and not significant in Weeks 9–17. These results provide very weak support for Hypothesis 3b. Again, the value of growth modeling is demonstrated by its ability to show the complexity of IT use over time: while the influence of age is largely through direct effects, the effect of gender is based on a combination of direct and rate of change effects.

Academic Performance (GPA)

Results from tests of Hypotheses 4a and 4b, which concern the associations between GPA and BIStro use/week over time, were interesting. The comparison of Models 1B and 1C showed no significant association between BIStro use and GPA during the first week of class (β_{03} = N/S). These results do not support Hypothesis 4a. A comparison of Model 1C and 1D did not show significant moderating effects between GPA and variations in changing rates of BIStro use (β_{13} = N/S).

The results differed in the second half of the class. Students with higher GPA exhibited higher rates of BIStro use during Week 9 (β_{03} = 0.048, p < 0.05 in Model 2C; 0.073, p < 0.01 in Model 2D). However, the rates of change in BIStro use over the second half of the class were not associated with GPA. These results support Hypothesis 4a, but in a way not reported in previous research (i.e., a direct effect was found halfway through the classes, but not at the beginning). This latter finding suggests that there are critical times in the semester (e.g., after a midterm exam) when IT use can be associated with academic success. The results do not support Hypothesis 4b.

Implications

Our results showed that growth modeling helped BIStro designers better understand their students' use of BIStro over the course of a semester. Results show that BIStro use is associated with personal characteristics such as gender or GPA in complex ways that vary over time. While the specific findings are limited to the BIStro system, they might make instructors think twice about taking a "blanket approach" to IT use throughout an entire semester, year, or program of study. Instead, instructors might need to adjust their strategies and tactics for encouraging IT use over the course of the semester.

Limitations

As in all research, our evaluation of BIStro does have limitations. First, the data for the tutorial was collected from only one IT implementation, so the results may not be widely generalizable to other systems. Second, the dependent variable used in the study, total number of log-ins per week, does not give a complete picture of the complexity of IT use. As noted previously, technology use is sporadic and decisions regarding acceptance and use do not occur at one point in time. Future researchers who are interested in this line of research should attempt to capture users' changing perceptions over time, as well as other aspects of IT use (e.g., length of IT use sessions and type of use). The use of autoregressive models may be particularly helpful in this regard. Third, data about important individual characteristics such as ethnicity were not available, which may explain the low amounts of variance explained. Future research should study the ways in which other event- and person-level variables are associated with technology acceptance and IT use over time [Davis et al. 2004]. For example, researchers may want to explore the effects of *academic competence*, which is "a multi-dimensional construct composed of the skills, attitudes, and behaviors of a learner that contribute to academic success" [DiPerna 2004: 64], rather than just past academic performance (i.e., GPA).

Several general issues with longitudinal research are relevant to the specific case of technology acceptance and use. At the most basic level, longitudinal data is often difficult to obtain, as organizations may be reluctant to provide access to individuals over time. Researchers must first be concerned, therefore, with securing appropriate sample frames for longitudinal research. Further, the analysis of longitudinal data and the interpretation of results must account for the complexity of the data and the potential confounds that are unique to longitudinal research (e.g., autoregression). Researchers must plan such research carefully and, to the extent possible, avoid contamination of data by eliminating or controlling for confounding variables.

VI. CONCLUSIONS

Time is an important factor in the use of information technology, but must be better incorporated into IS research. Results from our growth model analysis showed that user characteristics differ in the ways they influence the relationship between IT use and time: the influence of age and academic performance was based solely on direct effects while that of gender was based on a mix of direct and rate of change effects. Moreover, the results showed the strength of these effects can change between periods of time. The richness of the analysis provided by growth modeling methods helps us better understand the complex and dynamic relationships between time, information technology, and human behavior.

The tutorial demonstrates the usefulness of growth modeling theory and HLM methods in analyzing IT use over time. It describes growth modeling's conceptual flexibility and analytical power over traditional IS methods such as One-way ANOVA and OLS regression. We believe growth modeling has a bright future in IS research. Relationships among a variety of individual-, team-, group-, or higher-level factors remain to be explored, including how they might be associated with time and IT use. Other types of information systems such as customer relationship management, business intelligence, or more traditional implementations, such as group decision support systems, should be examined.

We hope this tutorial on HLM and growth modeling analysis can generate interest among IS researchers and help them appreciate the sophistication and usefulness of growth modeling in their research efforts. We want to emphasize that HLM is not the only method that can be used in growth modeling; for example, SEM can also be used. While each has their advantages, ongoing research may have cross-pollinating effects. As Bollen and Curran [2006] argue:

> *…there is an exciting trend toward the convergence of the SEM techniques and multilevel [HLM] techniques such that each approach stimulates the others' development. We anticipate that both approaches will benefit from the continuing interaction between practitioners of each* (p. 262).

We also believe growth modeling—and other types of multilevel modeling—can help IS researchers gain a deeper understanding of the intricacies of cross-level effects. The BIStro case provides several examples. First, the relationships between IT use and time, gender, age, and academic performance are more complex than previously reported in the literature. Age was found to be directly related to IT use, though this effect was found to be stronger in the first half of the semester than in the second. Gender was also found to be directly associated with IT use, but it also moderated the relationship between time and IT use during the first and second halves of the semester. These findings are important because age and gender are often used as control variables. However, unlike other research showing gender effects to be consistent over time, our study found that they can indeed change over time. Results from our tutorial show the methodological value of growth modeling over other alternatives that use "snapshots" of behavior or violate assumptions of traditional methodologies. We recommend its use in research about other temporal cross-level effects as well.

REFERENCES

Editor's Note: The following reference list contains hyperlinks to world wide web pages. Readers who have the ability to access the web directly from their word processor or are reading the paper on the web, can gain direct access to these linked references. Readers are warned, however, that

1. these links existed as of the date of publication but are not guaranteed to be working thereafter.
2. the contents of web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. the author(s) of the web pages, not AIS, is (are) responsible for the accuracy of their content.
4. the author(s) of this article, not AIS, is (are) responsible for the accuracy of the URL and version information.

Afshartous, D. (1995). "Determination of Sample Size for Multilevel Model Design," *Annual Meeting of the American Educational Research Association, San Francisco, CA*. Retrieved 2 September 2009 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.4318.

Agarwal, R., and V. Venkatesh (2002). "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability," *Information Systems Research* (13)2, pp. 168–186.

Aguinis, H., et al. (2009). "Customer-Centric Science: Reporting Significant Research Results with Rigor, Relevance, and Practical Impact in Mind," *Organizational Research Methods* (Forthcoming). Retrieved 8/25/2009 from http://orm.sagepub.com/cgi/rapidpdf/1094428109333339v1.

Ahuja, M.K., and J.B. Thatcher (2005). "Moving beyond Intentions and Toward the Theory of Trying: Effects of Work Environment and Gender on Post-Adoption Information Technology Use," *MIS Quarterly* (29)3, pp. 427–459.

Ancona, D.G., et al. (2001). "Time: A New Research Lens," *Academy of Management Review* (26)4, pp. 645–663.

Ang, S., S. Slaughter, and K.Y Ng (2002). "Human Capital and Institutional Determinants of Information Technology Compensation: Modeling Multilevel and Cross-Level Interactions," *Management Science* (48)11, pp. 1427–1445.

Benbasat, I., and H. Barki (2007). "*Quo vadis*, TAM?" *Journal of the Association for Information Systems* (8)4, Article 3, pp. 211–218.

Bickel, R. (2007). *Multilevel Analysis for Applied Research: It's Just Regression!* New York: Guilford Publications.

Bliese, P.D., and T.W. Britt (2001). "Social support, group consensus and stressor-strain relationships: Social context matters," *Journal of Organizational Behavior* (22)4, pp. 425–436.

Bliese, P.D., and R.E. Ployhart (2002). "Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations," *Organizational Research Methods* (5)4, pp. 362–387.

Bollen, K.A., and P.J. Curran (2004). "Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions," *Sociological Methods & Research* (32)3, pp. 336–383.

Bollen, K.A., and P.J. Curran (2006). *Latent Curve Models: A Structural Equation Perspective*. Hoboken, NJ: Wiley.

Brotherton, J.A., and G.D. Abowd (2004). "Lessons Learned from eClass: Assessing Automated Capture and Access in the Classroom," *ACM Transactions on Computer-Human Interaction* (11)2, pp. 121–155.

Bryk, A.S, V. E. Lee, and P.E. Holland (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Burton-Jones, A., and D.W. Straub (2005). "A Multilevel Investigation of the Relationship between System Usage and Task Performance." Working paper 05-MIS-018, University of British Columbia. Retrieved 23 March 2006 from http://mis.sauder.ubc.ca/research/workingpapers.html.

Cohen, J., et al. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum.

Davis, F., M.G. Morris, and V. Venkatesh (2004). "Technology Adoption Research: Does It Have a Future?" Invited panel discussion, *Decision Sciences Institute 2004 Meeting,* Boston, MA.

Deadrick, D.L., N. Bennett, and C.J. Russell (1997). "Using hierarchical linear modeling to examine dynamic performance criteria over time," *Journal of Management* (23)6, pp. 745–757.

Devaraj, S., and R. Kohli (2003). "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?" *Management Science* (49)3, pp. 273–289.

DiPerna, J.C. (2004). "Structural and Concurrent Validity Evidence for the Academic Competence Evaluation Scales—College Edition," *Journal of College Counseling* (7)1, pp. 64–72.

Enders, C.K., and D. Tofighi (2007). “Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue,” *Psychological Methods* (12)2, pp. 121–138.

Fallows, D. (2005). “How Women and Men Use the Internet,” Pew Internet & American Life Project, Washington, DC. Retrieved 9/21/2009 from http://www.pewinternet.org/Reports/2005/How-Women-and-Men-Use-the-Internet.aspx.

Gefen, D., and D.W. Straub (1997). "Gender Differences in the Perception and Use of E-mail: An Extension to the Technology Acceptance Model,” *MIS Quarterly* (21)4, pp. 389–400.

He, J., B.S. Butler, and W.R. King (2007). “Team Cognition: Development and Evolution in Software Project Teams,” *Journal of Management Information Systems* (24)2, pp. 261–292.

Hitt, M.A., et al. (2007). “Building Theoretical and Empirical Bridges Across Levels: Multilevel Research in Management,” *The Academy of Management Journal* (50)6, pp. 1385–1399.

Hofmann, D.A. (1997). “An Overview of the Logic and Rationale of Hierarchical Linear Models,” *Journal of Management* (23)6, pp. 723–744.

Hofmann, D.A. (2006). “Hierarchical Linear Modeling”, professional development workshop, *Academy of Management 2006 Annual Meeting*, 11–16 August, Atlanta.

Hofmann, D.A., and M.B. Gavin (1998). “Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations,” *Journal of Management* (24)5, pp. 623–641.

Hofmann, D.A., R. Jacobs, and J.E. Baratta (1993). “Dynamic Criteria and the Measurement of Change,” *Journal of Applied Psychology* (78)2, pp. 194–204.

Hourcade, J.P., et al. (2004). “Differences in Pointing Task Performance between Preschool Children and Adults Using Mice,” *ACM Transactions on Computer-Human Interaction* (11)4, pp. 357–386.

James, L.R., and L.J. Williams (2000). “The Cross-Level Operator in Regression, ANCOVA, and Contextual Analysis” in Klein, K. and S.W.J. Kozlowski (eds.). *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco: Jossey-Bass, pp. 382–424.

Klein, K.J., and S.W.J. Kozlowski (2000). *Multilevel Theory, Research, and Methods in Organizations.* San Francisco: Jossey-Bass.

Kozlowski, S.W.J., and K.A. Klein (2000). “Multilevel Approach to Theory and Research in Organizations: Contextual, Temporal, and Emergent Processes” in Klein, K., and S.W.J. Kozlowski (eds.). *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco: Jossey-Bass, pp. 3–90.

Lapointe, L., and S. Rivard (2005). “A Multilevel Model of Resistance to Information Technology Implementation,” *MIS Quarterly* (29)3, pp. 461–491.

Lee, H., et al. (2006). “User Evaluation of Físchlár-News: An Automatic Broadcast New Delivery System,” *ACM Transactions on Information Systems* (24)2, pp. 145–189.

Maas, C.J.M., and J.J. Hox (2004). “Robustness Issues in Multilevel Regression Analysis,” *Statistica Neerlandica* (58)2), pp. 127–137.

Maas, C.J.M., and J.J. Hox (2005). “Sufficient Sample Sizes for Multilevel Modeling,” *Methodology* (1)3, pp. 86–92. Retrieved 2 September 2009 from http://igitur-archive.library.uu.nl/fss/2007-1113-201005/UUindex.html.

Malcolm, S., D.E. Chubin, and E. Babco (2005). “Women and STEM Disciplines: Beyond the Barriers,” *On Campus with Women* (34)4, Association of American Colleges and Universities. Retrieved 27 February 2006 from http://www.aacu.org/ocww/volume34_4/feature.cfm?section=2.

Malhotra, Y., D.F. Galletta, and L.J. Kirsch (2008). “How Endogenous Motivations Influence User Intentions: Beyond the Dichotomy of Extrinsic and Intrinsic User Motivations,” *MIS Quarterly* (25)1, pp. 267–299.

Mithas, S., et al. (2006–7). “Designing Web Sites for Customer Loyalty Across Business Domains: A Multilevel Analysis,” *Journal of Management Information Systems* (23)3, pp. 97–127.

Murnane, R.J., K.P. Boudett, and J.B. Willett (1999). “Do Male Dropouts Benefit from Obtaining a GED, Postsecondary Education, and Training?” *Evaluation Review* 23, pp. 475–502.

Orlikowski, W.J., and C.S. Iacono (2001). “Research Commentary: Desperately Seeking the “IT” in IT Research—A Call to Theorizing the IT Artifact,” *Information Systems Research* (12)2, pp. 121–134.

Palmer, J.W. (2002). “Web Site Usability, Design, and Performance Metrics,” *Information Systems Research* (13)2, pp. 151–167.

Papadakis, M.C. (2000). “Complex Picture of Computer Use in the Home Emerges,” *Issue Brief*, National Science Foundation, Division of Science Resource Studies, March 31, 2000. Retrieved 28 February 2006 from http://www.nsf.gov/statistics/issuebrf/sib00314.pdf.

Petersen, M.G., K.H. Madsen, and A. Kjaer (2002). “The Usability of Everyday Technology—Emerging and Fading Opportunities,” *ACM Transactions on Computer-Human Interaction* (9)2, pp. 74–105.

Raudenbush, S.W. (2001). “Comparing Personal Trajectories and Drawing Causal Interferences from Longitudinal Data,” *Annual Review of Psychology* 52, pp. 501–525.

Raudenbush, S.W., and A.S. Bryk. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition. Newbury Park, CA: Sage.

Raudenbush, S., et al. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Raudenbush, S., A. Bryk, and R. Congdon (2008). *HLM for Windows, Version 6.06*. Scientific Software International. Software available 12/13/2008 from http://www.ssicentral.com/.

Scherbaum, C.A., and J.M. Ferreter (2009). “Estimating Statistical Power and Required Sample Sizes for Organizational Research Using Multilevel Modeling,” *Organizational Research Methods* (12)2, pp. 347–367.

Scientific Software International (2008). Basic Examples. Retrieved 12/7/2008 from http://www.ssicentral .com/hlm/examples.html.

Singer, J.D., and J.B. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Snijders, T. (2005). “Power and Sample Size in Multilevel Modeling,” in B.S. Everitt and D.C. Howell (eds.) *Encyclopedia of Statistics in Behavioral Science* (3), pp. 1570–1573. Retrieved 2 September 2009 from http://stat.gamma.rug.nl/PowerSampleSizeMultilevel.pdf.

Snijders, T., and R. Bosker (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.

Subrahmanyam, K., et al. (2000). “The Impact of Home Computer Use on Children’s Activities and Development,” *The Future of Children: Children and Computer Technology* (10)2, pp. 123–144.

The University of Texas at Austin, Division of Statistics and Scientific Consulting (2008). “Getting started with HLM for Windows.” Retrieved 19 November 2008 from http://ssc.utexas.edu/consulting/tutorials/stat/hlm/.

U.S. Department of Commerce, National Telecommunications and Information Administration (2004). “A Nation Online: Entering the Broadband Age.” Retrieved 21 September 2009 from http://www.ntia.doc.gov/reports/ anol/NationOnlineBroadband04.htm.

Venkatesh, V., and M. Morris (2000). “Why Don’t Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior,” *MIS Quarterly* (24)1, pp. 115–139.

Venkatesh, V., et al. (2003). “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly* (27)3, pp. 425–478.

von Bertalanffy, L. (1968). *General Systems Theory: Foundations, Development, Applications*. New York: Braziller.

APPENDIX

A checklist for researchers and reviewers.

Item	References
1. Is the selected growth modeling procedure appropriate?	
a. If the data set has randomly missing data, select hierarchical multivariate linear modeling as the MDM type (i.e., the HMLM or HMLM2 selections in Figure 3 of this paper). If the data set has no missing data, then hierarchical linear models can be used (i.e., the HLM2 or HLM3 selections in Figure 3).	Raudenbush et al. 2004: 140–148.
b. If measurement error must be accounted for, LGM is preferable over HLM.	Singer and Willett 2003: 280–295; Bollen and Curran 2006.
c. If the dependent variable will be predicted at many different times, a fixed- or variable-occasion model may be preferable to growth models.	Snijders and Bosker 1999: 167–180.
2. Is the sample size sufficient? Are there enough individuals (i.e., level-2 entities)?	
a. If level-1 effects are of primary interest, level-1 sample size is most important; if level-2 effects are of primary interest, level-2 sample size is most important.	Snijders 2005: 2.
b. If research interests require accurate, reliable variance components, "a relative large number" of level-2 units are necessary"; if interests require fixed effects estimates, "the number of level-2 units "decreases substantially."	Afshartous 1995: 12.
c. Effect size influences level-1 and level-2 sample size requirements.	Scherbaum and Ferreter 2009; see Figures 1–3, pp. 358–360, for estimates of statistical power for varying effect and sample sizes.
d. Small sample sizes can badly bias standard errors and subsequent statistical tests.	Maas and Hox, 2004: 135; Maas and Hox, 2005.
e. Highly accurate level-2 variance estimates may require 100 or more level-2 units (e.g., individuals in growth modeling).	Maas and Hox, 2004: 128.
3. Does the data collection match the theoretical entrainment processes?	
a. "At some points in the [process] cycle, two entities or levels may be tightly coupled or *entrained* [emphasis added], whereas at other points they will be decoupled and will appear independent." That is, ensure that data collection timing is consistent with underlying theory.	Kozlowski and Klein 2000: 24.
4. Is the data structure suitable for growth modeling?	
a. Data structures for growth modeling should take a person-period format rather than a person-level format. "In a person-period format, also known as *univariate format*, each individual has multiple records, one for each period in which he or she was observed" [Singer and Willett 2003: 22].	Singer and Willett 2003: 16–23.

5. Are key assumptions of random coefficients modeling tenable?	
a. "Each r_{ij} is independent and normally distributed with a mean of 0 and variance σ^2 for every level-1 unit *i* within each level-2 unit *j*." • The homogeneity of level-1 residuals (i.e., homoscedasticity) can be tested in the HLM software. See [Raudenbush et al., 2004: 59–60] for directions. • Level-1 residuals can also be examined in other ways. Normality and independence can be examined using other statistical packages (e.g., Q-Q plots, stem-and-leaf plots, or correlation matrices in SPSS). These tests require the generation of a level-1 residual file, directions for which can be found at [Raudenbush et al. 2004: 36-39].	Assumptions are quoted from Raudenbush and Bryk 2002: 255. Recall that in growth modeling, the level-1 residual is represented by *e* (e.g., e_{ti} in Equation 5a); the level-2 residual, by *r* (e.g., r_{0i} in Equation 5b). See also Raudenbush and Bryk 2002: 252–287; Singer and Willett 2003: 127–132; Snijders and Bosker 1999: 120–139; and Raudenbush et al. 2004.
b. "The level-1 predictors, X_{qij}, are independent of r_{ij}." • Compare level-1 residuals. The level-1 residual file described in Item 5.a above will also contain level-1 and level-2 predictor values. Independence can be tested using other statistical packages (e.g., SPSS).	
c. "The vectors of Q + 1 random errors at level 2 are multi-variate normal, each with a mean of 0 some variance τ_{qq}, and covariance among the random elements, q and q', of $\tau_{qq'}$, The random-error vectors are independent among the J level-2 units." • Examine level-2 residuals. Directions for generating a level-2 residual file can be found at [Raudenbush et al., 2004: 39–47]. Multivariate normality and covariance can be examined using other statistical packages.	
d. "The set of level-2 predictors (i.e., all the unique elements in W_{sj} across the Q + 1 equations) are independent of every u_{qj}." • Compare level-2 predictors and level-2 residuals. The level-2 residual file described in Item 5.c above will also contain level-2 predictor values. Independence can be examined using other statistical packages (e.g., SPSS).	
e. "The errors at level-1 and level-2 are also independent." • Compare level-1 and level-2 residuals. Independence of residuals can be examined using other statistical packages (e.g., SPSS).	
f. "The predictors at each level are not correlated with the random effects at the other level." • Compare level-1 predictors and level-2 predictors, and level-2 predictors and level-1 residuals. Correlation matrices and residual plots can be generated using other statistical packages (e.g., SPSS).	
6. Does the intraclass correlation coefficient suggest the existence of sufficient level-2 variance to justify growth modeling?	
a. Theory should guide threshold levels of level-2 variance.	Raudenbush and Bryk, 2002: 36; Singer and Willett 2003: 96.
7. Has the data been tested for linearity?	
a. Non-linear trajectories can be transformed into linear trajectories.	Singer and Willett 2003: 210–213.
b. Non-linear trajectories can be divided into a series of linear "epochs."	Raudenbush and Bryk 2002: 178; Singer and Willett 2003: 206, 233.

8. Do the structural changes in succeeding growth models match the structure of the hypotheses? a. "variance explained in a level-2 parameter…is conditional on a fixed level-1 specification," so level-1 models should be developed first, with level-2 variables added afterwards.	 Raudenbush and Bryk 2002: 150.
9. Is the centering of variables reasonable and justified given the study's goals? Does the centering method match the paradigm of the study? a. Under the incremental paradigm (i.e., where "group level variables act as main effects in the prediction of individual-level outcomes"; [Hofmann and Gavin, 1998: 634]), grand mean centering is appropriate for level-1 variables. Group mean centering is also appropriate if means are added back to level-2 intercept model. b. Under the moderational paradigm (i.e., where "group level variables moderate the relationships between two individual-level variables"; [Hofmann and Gavin, 1998: 636]), grand mean centering can confound cross-level and between-group interactions. Group mean centering can be used to differentiate and check cross-level *versus* between-group effects.	 Hofmann and Gavin 1998: 634. Hofmann and Gavin 1998: 632–633, 636–637.
10. Is the selected maximum likelihood estimation procedure appropriate (i.e., full *versus* restricted)? For example, is the selected maximum likelihood estimation procedure consistent with the increase in the number of parameters across models? a. "…if you have applied [full maximum likelihood] estimation…you can use deviance statistics to test hypotheses about any combination of parameters, fixed effects, or variance components. But if you have used [restricted maximum likelihood] to fit the model, you can use deviance statistics to test hypotheses only about variance components…. Before using deviance statistics to test hypotheses, be sure you are clear about which method of estimation you have used" [Singer and Willett, 2003: 118]. b. "…using [full maximum likelihood], any pair of nested models can be tested using a likelihood ratio test. In contrast, using [restricted maximum likelihood], the likelihood ratio test is available only for testing variance-covariance parameters" [Raudenbush et al., 2004: 11; see also Table 1.1, page 12].	 Singer and Willett 2003: 117–119. Raudenbush et al. 2004: 11.

11. Has the analysis been interpreted reasonably? a. Is there a reasonable number of models in the analysis? • "When writing up findings for presentation and publication, we suggest that you identify a manageable subset of models that, taken together, tells a persuasive story parsimoniously." b. Are model comparisons sound? • The same data should be used in subsequent models. • Previous models should be nested in subsequent models. c. How well does the addition of new variables improve a previous model? • Variables added to a previous model may be significant, though model itself may not be improved significantly. Model improvement should be checked with a likelihood ratio test (i.e., model χ^2 difference test). d. Is the use of pseudo-R^2 statistics appropriate? • There is wide disagreement on the value of such statistics, and anomalies do occur. Researchers and reviewers should thoughtfully consider the value and limitations of these statistics. e. Should a "final" model be included? • A "final" parsimonious model allows the researcher to check changes in parameters when non-significant variables are deleted.	Singer and Willett 2003: 106. Singer and Willett 2003: 118. Singer and Willett 2003: 118. Singer and Willett 2003: 116–120. Singer and Willett 2003: 102–104; Snijders and Bosker 1999: 99–105. Singer and Willett 2003: 109–110
12. Should random or fixed coefficients be used? a. If the researcher is interested in level-2 effects (e.g., the influence of individual-level characteristics in longitudinal studies), then random models will provide the requisite analysis. On the other hand, if level-2 effects are not the focus of the study, or if their addition would unnecessarily complicate the model, then a fixed model would be the simpler, more parsimonious choice.	Raudenbush and Bryk 2002: 135–139

ABOUT THE AUTHORS

Robert F. Otondo is an Associate Professor of Information Systems at Mississippi State University. His research interests center on the perceptions and use of emerging technologies. His research has been published in *Information & Management*, *Journal of Applied Psychology*, *Production and Operations Management*, *Decision Support Systems*, and *Communications of the Association for Information Systems*. His research has been funded by the Office of Naval Research, and by a grant from the Robert Wood Johnson Foundation through the FedEx Center for Supply Chain Management at The University of Memphis.

Tim Barnett is currently a Professor of Management and the Richard and Mary Puckett Notable Scholar at Mississippi State University. He serves as the Coordinator of the Management Ph.D. program and Area Coordinator of the Interdisciplinary Leadership Minor. His research has appeared in the *Academy of Management Journal*, *Personnel Psychology*, *Human Relations*, *Entrepreneurship Theory & Practice*, *Business Ethics Quarterly*, *Journal of Business Ethics*, and the *Journal of Business Research*, among other journals. Tim serves on the editorial boards of the *Academy of Management Journal*, the *Journal of Management*, and *Family Business Review*.

Franz W. Kellermanns is the Henry Family Notable Scholar and Associate Professor of Management in the College of Business and Industry at the Mississippi State University. He holds a joint appointment with the INTES Center for Family Enterprises at the WHU-Otto Beisheim School of Management (Germany). He received his Ph.D. from the University of Connecticut. His research interests include strategy process and entrepreneurship with a focus on family business research. He is an Associate Editor of *Family Business Review* and has published in journals such as the *Journal of Management*, *Journal of Management Studies*, *Journal of Organizational Behavior*, *Journal of Business Venturing*, *Entrepreneurship Theory and Practice*, *Family Business Review*, and *Academy of Management Learning and Education*.

Allison W. Pearson is a Professor of Management at Mississippi State University. Her research interests include cognitive modeling and attitude formation. Her research has been published in *Journal of Management*, *Small Group Research*, *Entrepreneurship Theory and Practice*, *Journal of Applied Social Psychology*, and *Journal of Computer Information Systems*. Dr. Pearson has worked with many professionals through outreach programs funded by grants from the US Department of Education.

Rodney A. Pearson is a Professor of Information Systems at Mississippi State University. His research in computer-assisted learning has been published in *e-Service Journal*, the *Annual Conference of the Association for Business Communication*, and the *Annual Conference of the Southern Management Association*. Dr. Pearson has participated in several outreach programs.

Communications of the Association for Information Systems

ISSN: 1529-3181