A Methodological Primer for Conducting Quantitative Research in Postsecondary Education at Lumina Foundation for Education

Fred J. Galloway
University of San Diego

January 2004

Introduction

For the last thirty years, the rapid growth in information technologies has led to an absolute explosion in the production and diffusion of statistical software. With drop-down menus replacing meticulously written computer code, it now seems that literally *anyone* can run a three-way analysis of variance, perform generalized statistical tests for a proportional hazards model, or use varimax rotation to maximize the value of factor loadings. Despite the fact that a nuanced understanding of techniques like these takes years of theoretical study and empirical practice, aspiring quantitative researchers can now perform these procedures in a matter of minutes on such popular types of statistical software as SAS and SPSS. As a result, there has been a growing mismatch between what many quantitative researchers are *able to do* and what they are *able to understand.*

This growing mismatch has already had profound implications on many disciplines in the social sciences traditionally more comfortable with the tools of qualitative research. Although the gap between knowledge and practice has been felt in such social scientific disciplines as political science, anthropology, and sociology, nowhere has the effect been more pronounced than in the field of educational research, which has traditionally produced some of the most poorly designed quantitative research studies of all. In fact, the state of quantitative research in the profession was so bad in 1999 that it led Bruce Thompson in his invited American Educational Research Association address to focus on the "constellation of seven cardinal sins of analytic research practice" (Thompson, 1999).

In an effort to bridge this rapidly growing methodological divide, this paper will help prospective grantees select the most appropriate set of analytical techniques for their research questions so that the results are as inferentially robust and generalizable as possible. This will be accomplished by first discussing the different types of databases available to quantitative researchers and the five types of research questions that these databases can support. For each of these research questions, the appropriate set of analytical techniques will then be discussed so that researchers will learn how to efficiently match techniques to questions. This will be followed by a discussion of the various data problems that quantitative researchers can expect to encounter and their respective solutions. The paper then concludes with a section on how the results of these models can be presented in a clear and insightful manner, so that after readers have learned *what* to present, in this final section they can learn *how* to present it.

Research Question #1: Conceptually, what are the different types of databases available for quantitative research in higher education and what sorts of research questions can they support?

To make sense of the different types of databases available for quantitative research in higher education, it's helpful to recall the difference between a population and a sample, since databases can appear in either form. When a database contains *every* single element in a group, like all of the students admitted to a school, or all the student loans held by a particular organization, then the database is said to contain the entire population. Examples of databases like this in postsecondary education are typically institutional -- like all those students admitted to the University of San Diego for academic year 2002-03, or all of those students whose loans are currently held by Sallie Mae. Databases like this can be contrasted with those that contain only samples taken from the population, like the National Postsecondary Student Aid Survey (NPSAS), which samples both schools and students every five years, or the National Longitudinal Study of Youth, which sampled 12,686 students between the ages of 14 and 22 in 1979.

Since databases that contain samples drawn from populations are the most common type of databases in education, a further distinction needs to be made in how the samples are drawn. With *probability samples*, the probability that each member, or element of the population, is included in the sample is known by the researcher, where in *judgment samples*, personal judgment plays a major role in determining which elements of the population are selected, and this probability is not known. For example, if a list of every Historically Black College and University (HBCU) were compiled and every other one selected, then this would be a probability sample, with the probability of selection given by .5. However, if another sample was selected based solely on whether the researcher had personally visited a particular HBCU, then this would be a judgment sample and the probability of selection would be unknown.

The distinction between probability and judgment samples is an important one because researchers working with probability samples have the ability to estimate how large their sampling error is, however with judgment samples there is no way to tell how "far off' a sample result is likely to be. For this reason, most databases are constructed from probability samples, and estimates of the sampling error provided, making it easier to discuss how much confidence the researcher has that the results they found in the sample also occur in the population. This ability to generalize from the sample to the population is perhaps the most fundamental and necessary condition for any well-designed quantitative study, and all researchers should be willing to address this issue in the methodology section of their proposal. Since this condition can never be met by judgment samples, they are to be avoided unless both financial and logistical concerns necessitate their use.

Although there are many types of probability samples, including stratified random samples, systematic samples, and cluster samples, all of them are characterized by an element of randomness in the actual selection of the sample. Whether the randomness appears as selecting every kth element on a list of all elements in the population, or simply dividing a population into strata and then taking a random sample from each of the strata, the notion of random selection is what probability samples are all about. And since the probability of selection in probability samples is always known, researchers have the ability to "weight up" the responses from individuals in the sample to represent the entire population on such key dimensions as gender, race/ethnicity, income, and location.

To illustrate both the power and importance of weighting up sample data to reflect the population, consider a population of 1000 undergraduates at a small liberal arts college – 400 men and 600 women. Now suppose that the names of all 1000 undergraduates are placed on a list alphabetically and every other person is selected for a sample of size 500. Suppose further that only 400 of the 500 undergraduates respond, and that 300 of the respondents are women and only 100 are men. Of course, if you just report the results from your sample, you'd be reporting on what 300 women and 100 men had to say, despite the fact that in the population there were more men proportionally (40% of the population yet only 25% of the responding sample), The way to solve this problem, in other words the way to make the sample truly represent the population in terms of gender is to use the weighting concept, where each individual simply has a weight given by the inverse of their probability of selection. In this example, since there were 100 men in the responding sample and 400 men in the population, each man gets a weight of 4, and since there were 300 responding women and 600 in the population, each woman gets a weight of 2. These weights allow researchers to "scale up" the sample so that it matches the population; in this example the weighting was done by gender so that the weighted sample matches the population exactly in terms of gender.[1] As mentioned earlier, the weights in this example represent the inverse of the probability of selection, since the probability of selection for men was 100/400 or .25 and the weight 1/. 25 = 4; for the women the probability of selection was 300/600 or .5 and the weight was 1/. 5 = 2.

Although the above example may seem a bit long and tedious, the notion that weighting can be used to match samples demographically to their respective populations is a powerful analytic tool that can help researchers make more informed judgments about how a population behaves. Since this sort of weighting is impossible without knowing the probability of selection, probability samples are to be used wherever possible. Of course, there may be times when it is impractical or impossible to completely describe a particular population, as might be the case when studying a Tribal College that can't provide an accurate enrollment count because of insufficient or outdated records, or a proprietary school that might not wish to disclose enrollment information for a variety of reasons. Although such exceptions do occur, whenever possible, quantitative research proposals should be based on databases formed from entire populations or probability samples drawn from populations.

From this class of admissible databases, a further distinction can be made based on whether the researcher is looking at things at one point in time or looking at things over time. If the former is true, the database is said to be *cross-sectional* in that the research focus is on looking at things at one particular point in time, while if the later is true then the database is said to be *time-series* in that it is looking at how things have changed over time. If both conditions are true, the database is said to be *panel* or *longitudinal*, in that it may contain observations on thousands of individuals, each observed at numerous points in time. For example, the 1993-94 NPSAS database is cross-sectional in that it looks at sources of student financial support at one point in time, that being academic year1993-94, but when the degree completers in this database were then followed over time, the resulting database, Baccalauriate & Beyond, became longitudinal. This can be contrasted with a pure time-series database such as the number of undergraduates admitted each year to Harvard College over the last century, where one can plainly see that the same individuals are not being followed over time. Although all three types of databases have their place in empirical research, the longitudinal databases have the greatest potential for powerful inferences, since they provide variation both at individual points in time, as well as over time.

Of course, no matter what type of database is used in the analysis, researchers need to explicitly specify their research questions and choice of analytical techniques in their proposal. Since the choice of an appropriate set of analytical techniques is key to any meaningful analysis, and since there are an increasing number of possible techniques to choose from, the matching of techniques to research questions is one of the most important, yet vexing decisions faced by quantitative researchers. In an effort to make this matching decision as easy and formulaic as possible, this paper will use a classification system that allows all research questions to be reduced to just five types – those involving the degree of relationship among variables; those involving the significance of group differences; those involving the prediction of group membership; those involving the search for some underlying structure; and those involving the time course of events (Tabachnick and Fidell, 2001). This typology will be used in the next section as a method for organizing the array of possible analytical techniques.

Research Question #2: For each type of research question, what analytical techniques provide the most robust set of inferences and how can these models be estimated?

Since the databases discussed in the first section support five sorts of research questions, in this section the appropriate set of analytical techniques for each type of research question will be discussed. With this information, potential grantees will be able to select the most inferentially robust set of techniques for their particular set of research questions. More importantly, this typology will show that when researchers carefully match their techniques to their research questions, quantitative research can be produced that is both methodologically sound and generalizable, helping to inform the work of state, federal, and institutional policymakers.

The Degree of Relationship Among Variables

The first type of research question that will be considered is also the most popular – those questions involving the degree of relationship among variables. The discussion will begin with bivariate and multivariate correlation, followed by the techniques of multiple regression analysis and hierarchical multiple regression analysis.[2] Throughout the discussion, the focus will be on the methodological insights necessary to produce inferentially meaningful work, and should not be viewed as a substitute for a basic understanding of the underlying mathematical statistics.

In assessing the degree of relationship among variables, the most fundamental tool is the bivariate *correlation coefficient*, which measures the linear association between any two variables. This measure is unit-free and bounded between –1 and +1, so that a correlation close to +1 means that the two variables tend to move in the same direction together (i.e. both variables are either above their means, or below their means), while a correlation close to –1 means that the variables tend to move in opposite directions (one is below its mean while the other above its mean). In a similar manner, the multivariate correlation coefficient measures the association between a dependent variable and an optimally weighted combination of two or more independent variables.[3] However, unlike the simple bivariate correlation coefficient, the multivariate correlation coefficient is only bounded between 0 and +1, where a value of zero means that no relationship exists with the independent variables and a value of +1 indicating a perfect relationship.

Although correlation coefficients are helpful in assessing the degree of relationship among variables, they are less helpful when it comes to predicting values for the dependent variable or estimating the relative contribution of each of the independent variables in predicting the dependent variable. Fortunately, *multiple regression analysis* solves these problems by allowing researchers to select, or specify, a set of independent variables that the researcher believes may help explain why a particular dependent variable behaves in the way that it does. For example, multiple regression analysis might be used to explain why some colleges have higher graduation rates than others, or why some students have higher grade point averages than others.

In developing these multiple regression models, researchers must be extremely careful in their selection of independent variables since the omission of any statistically relevant variable typically leads to biased estimates of the contribution that each independent variable makes to the dependent variable.[4] This problem, called “specification error”, is one of the most pervasive problems in regression analysis and often times occurs when researchers are forced to select their variables from data already gathered by others. To see how a problem like this might occur, consider the example of why some colleges have higher graduation rates than others and imagine a thoughtful researcher trying to decide what independent variables to include in their model. One can easily imagine selecting at least three sorts of variables -- those relating to the quality and preparation of the student body, those describing the way that institution behaves towards students, and finally those variables that describe the net price that students pay at their respective college.[5] Although the information for the first two types of variables might be readily available from a database like the Integrated Postsecondary Education Data System (IPEDS) or from institutional records, the third bit of information might be very difficult, or even impossible to locate, since many financial aid offices may not keep data in this form. Of course when this occurs, the researcher has no choice but to limit their modeling effort to including only the first two types of variables and acknowledge that their model may suffer from specification error.

Although specification error is one of the most common problems in regression analysis, there is a simple two-step method for dealing with it. First, when actually specifying the variables to be used in the model, let theory be your guide in identifying what sorts of effects one might expect, and then make every effort to include all of the variables that represent these effects. Second, when data limitations prevent you from including all of the candidate variables, be sure to state explicitly that several variables that might have proven significant were not included in your analysis, and to the extent that these missing variables are correlated with any of the variables in the model, then your estimated coefficients may be biased. However, the specification of a model's independent variables is also the place where, as magicians say, "the rabbit goes into the hat" in that the effects or variables that appear significant in a model are largely a function of the variables that were initially included for testing in the model. In fact, in his 1983 article, "Let's Take the Con Out of Econometrics", Edward Leamer argues that through judicious selection of a model's candidate variables, researchers with very different backgrounds were all able to claim that their particular hypotheses are supported, despite the fact that in many cases these hypotheses were mutually exclusive. The bottom line is that unless researchers are both willing and able to discuss the *inferential fragility* of their results, in other words, how sensitive their results are to the inclusion of other variables, their results should simply not be believed or accepted.

This notion of inferential fragility is indeed a powerful concept that applies not only to how models are specified, but also to the size of the effects produced by the variables in a regression model. Although many researchers spend lots of time talking about how statistically significant some of their variables are, the more interesting and relevant question is what sort of effect is produced by each of the significant variables in a particular model. Although this topic will be discussed further in the context of the paper's fourth research question, the notion of inferential robustness suggests that instead of just producing a point estimate of the effect size for each of the variables, a range of effect sizes that are associated with slightly different specifications of the regression model may be more appropriate. For example, if there are three potential ways of measuring income in a particular regression model, the best way to gauge the size of the effect that *other* variables in the model may have on the dependent variable is to run three different regression models, each one containing a different measure of income. As the different models are run, of course, different effects are produced for all of the model's variables, and a range can be established that contains all three of the estimated effect sizes for a particular variable. As expected, the narrower the range for the effect sizes, the more stable, or robust, is the inference, while a wider range for the effect sizes suggests that the inference is more fragile and less likely to be believed.

By paying attention to the problems of specification error and inferential fragility, aspiring researchers can use the tools of multiple regression analysis to produce methodologically sound quantitative research. However, when attempting to disentangle the amount of variation in the dependent variable explained by the various types of independent variables in a particular model, *hierarchical multiple regression* analysis is more appropriate. In this type of analysis, variables are loosely grouped into several general categories and the contribution of each category of variables is then estimated by running a series of nested models, with each successive model adding another category of variables. For example, in our earlier example of why some colleges have higher graduation rates than others, recall that three sorts of variables were selected for analysis -- those relating to the quality and preparation of the student body, those describing the way that institutions behave towards students, and finally those variables that describe the net price that students pay at their respective college. Through the use of hierarchical multiple regression analysis, the contribution that each one of these three groups makes in helping to explain variation in graduation rates can be assessed by first regressing graduation rates against those variables relating to the quality and preparation of the student body, then running a second model where graduation rates are regressed against variables relating to the quality and preparation of the student body *and* variables describing the way that institutions behave towards students, and finally, running a third model where all three categories of variables are included. Methodologically, this allows the researcher to net out the effects of each of the three types of variables, so that the contribution that each of these categories makes can be easily estimated. And of course this can be critically important for policy matters, since if the quality and preparation of the student body explains twice as much of the variation in graduation rates as does the way that institutions behave toward students, then increasing graduation rates can be as simple as increasing admission standards, rather than concentrating on ways in which the institution can provide more for their current students. For this reason, it's not surprising that hierarchical multiple regression analysis has become an increasingly popular tool for serious quantitative researchers.

The Significance of Group Differences

The second type of research question involves the significance of group differences, and this discussion will cover the popular techniques of analysis of variance and analysis of covariance, both in the case of a single and multiple dependent variables. However, before examining the techniques used to test for the significance of group differences, an important distinction needs to be made between truly randomized experiments, where participants are randomly assigned to treatment and control groups, and non-randomized experiments, where subjects, for whatever reason, cannot be randomly assigned to these groups. This distinction is important because causality *can only* be assigned when randomized experiments are conducted, and in these situations, the popular techniques of analysis of variance and analysis of covariance are often used. For example, the researcher that randomly assigns children to either the Head Start program or some other control program may be able to argue that participation in the Head Start program caused children to score higher on some measure of early childhood achievement, while the researcher who simply observed that children who had participated in the Head Start program scored higher than children who did not cannot make this sort of causal inference concerning the effect of the popular program.

When the techniques of *analysis of variance* (ANOVA) are used, researchers are typically interested in testing for differences between groups in terms of their mean scores on some outcome measure. These techniques all revolve around looking at the variance of scores, both within groups and between groups. Specifically, ANOVA works by statistically comparing the within-group variance, which can be thought of as random or error variance, to the between-group variance, which can be thought of as a reflection of group differences. If these two variances are fairly similar, then one can conclude that all of the group means came from the same sampling distribution and the slight differences among them are due to random error. However, if the group means differ significantly, then the researcher can conclude that these means were drawn from different sampling distributions, and the null hypothesis that the means are the same can be rejected. Computationally, these tests of significance are done through the use of an F-test, developed by R.A. Fischer in the early 1920s for use with the techniques of ANOVA.

Although these techniques are widely used in the educational research literature, they do have significant limitations and are often used incorrectly. For example, researchers often use ANOVA techniques with non-randomized experiments, when the techniques of regression analysis would be more appropriate. Similarly, when it makes more sense to control for certain effects even within a randomized experiment, researchers still use ANOVA techniques when more advanced techniques like the analysis of covariance are more appropriate. And finally, if the fairly technical assumptions of ANOVA are not met, then the use of ANOVA may be at best, misleading, and at worst, factually inaccurate.[6]

Given their popularity, the techniques of ANOVA have been extended in many important ways. Perhaps the most popular extension has been the introduction of the *analysis of covariance*, commonly referred to as ANCOVA, which tests for differences in mean scores between groups *after* the scores have been adjusted for differences associated with one or more covariates (in the Head Start example these covariates might be the child's age and their parent's level of education). As with ANOVA, the variance in scores is divided into two groups – the variances of scores within groups and the variance in scores between groups – and essentially the same computation techniques utilized as with ANOVA. However, the techniques of ANCOVA are in many ways more general than those of ANOVA since these techniques can also be used with non-experimental data in a sort of "matching" function, where group scores can be adjusted to reflect what they would be if all of the subjects had identical values for the covariates. In addition, using ANCOVA also increases the power of the traditional F-test since it reduces the error term by accounting for some of the variance associated with the covariates. As such, unless random assignment has been performed and there are absolutely no candidate variables for covariates, the techniques of ANCOVA, *not* ANOVA should be used.

The techniques of ANOVA and ANCOVA have also been extended to the case of more than one dependent variable, and these parallel techniques are called the *multivariate analysis of variance* (MANOVA) and the *multivariate analysis of covariance* (MANCOVA). Although there are several advantages to using these techniques, perhaps the most important one is that by measuring several dependent variables instead of just one, researchers improve the chance of discovering exactly which outcome measure changes when the treatment changes. For example, the researcher studying the effects of the Head Start program might also want to examine the incidence of behavior problems among the children or perhaps their physical growth while in the program as well as how they scored on a test of early childhood achievement.

To get a feel for how MANOVA and MANCOVA work, recall that in ANOVA and ANCOVA researchers are interested in whether any mean differences by group have occurred by chance. In their multivariate extensions, researchers are interested in whether mean differences by group on *a combination of dependent variables* have occurred by chance. In the first stage of both MANOVA and MANCOVA, a new dependent variable is created from a linear combination of all the dependent variables that by construction, maximizes group differences, and then in the second stage a traditional ANOVA or ANCOVA is applied to this newly constructed dependent variable. And just as with their single dependent variable counterparts, the variance surrounding this newly constructed dependent variable can be broken down into the variances of scores within groups and the variance in scores between groups.

However, despite the fact that the techniques of MANOVA and MANCOVA are fairly straightforward from a computational standpoint, they require more stringent assumptions than ANOVA and ANCOVA and can be much more difficult to interpret statistically speaking. As such, they should only be used when the experimental design calls for more than one dependent variables, and even then, by researchers who have really paid their dues in multivariate statistics.

The Prediction of Group Membership

Although used less frequently in educational research than the first two types of research questions, the need occasionally arises for predicting group membership from a set of independent variables. Examples might include those students that receive institutional aid versus those that do not, or those former students who default on their student loans versus those who stay in repayment. In selecting the optimal analytical technique to use for the prediction of group membership, the choice essentially boils down to how the independent variables are measured -- specifically whether they are continuous, discrete, or a mixture of continuous and discrete. As such, this section will discuss three new techniques – discriminant function analysis, logistical regression analysis and multiway frequency analysis.

When all of the independent variables are continuous and normally distributed, *discriminant function analysis* is the preferred analytical technique for predicting membership in groups. Mathematically, this technique is equivalent to the techniques of MANOVA and conceptually can be thought of as “the flip side of the MANOVA coin”. To see this equivalence, recall that in MANOVA, group membership serves as the independent variable in the search for differences in a weighted combination of dependent variables, while in discriminant function analysis a weighted combination of independent variables are used to predict group membership, which serves as the model’s dependent variable. In fact, if groups differ significantly on a set of variables in MANOVA, the same set of variables can be used to reliably predict group membership in discriminant function analysis. However, because this type of analysis is typically used to predict membership in naturally occurring groups, rather than in groups formed by random assignment, important questions such as why we can reliably predict group membership or what causes differential membership are often not asked. In addition, researchers need to pay careful attention to the pattern of differences among the predictors as a whole to understand the dimensions along which groups differ.

However when all of the independent variables are discrete, multiway frequency analysis can be used to predict group membership. The purpose of this set of techniques is to discover associations among the independent variables by looking for relationships that help minimize the difference between observed cell frequencies and expected cell frequencies. The search for these associations begins with the highest order associations, and then works its way down to the one-way associations looking for statistical significance, so that in a model with three independent variables, the search would begin with the three-way association and then move on to all of the two-way associations, followed by the one-way associations. Although this set of techniques is used rather infrequently in the literature, it can be useful for examining research questions with discrete variables that involve the strength of associations among variables.

Last but certainly not least, when the independent variables are a mix of continuous and discrete, the optimal technique for predicting group membership is *logistic regression analysis*, a hybrid of standard regression analysis and logit analysis. Over the past two decades, this popular technique has been adopted by numerous social scientific disciplines, including educational researchers who have found it less restrictive than other techniques. For example, unlike discriminant function analysis and MANOVA, the independent variables do not have to be normally distributed, linearly related, or of equal variance within groups. Similarly, unlike multiway frequency analysis, the independent variables do not all have to be discrete; instead they can be a mix of continuous and discrete, and unlike standard regression analysis, logistic regression cannot produce negative predicted probabilities.

In addition to these advantages, logistic regression analysis also has the ability to handle more than two membership groups, and can model certain non-linear effects more efficiently than traditional regression analysis. However, interpreting the model's coefficients is much trickier than in standard regression analysis because there are three algebraically equivalent forms of the logistic regression equation, and each one has a different interpretation. For example, in an example described by Cohen, Cohen, West, & Aiken (2003), researchers interested in predicting the probability that assistant professors would be promoted to associate professor based on the number of publications they had would have three separate predictions would emerge from the logistic regression model – the logit, the odds of being promoted, and the predicted probability of being promoted. Although all three predictions can be derived from each other, the natural starting place is to begin with the logit, which looks like a typical regression in that it is linear in the coefficients but predicts a score in the unfamiliar log odds, or logit form. After the value of the logit has been calculated for *each* value that the independent variable assumes, then the odds associated with each level of publications can be calculated by simply taking the antilog of each logit. The final step is to then calculate the predicted probability associated with the different values of the independent variable from the odds of promotion. Since these calculations can be extremely tricky, researchers are urged to proceed with caution. In fact, since many of the traditional diagnostic measures of regression analysis are inappropriate for logistic regression, researchers interested in using these techniques are advised to spend some time studying the theoretical foundations for this sort of analysis before attempting to use this popular, but often misunderstood analytical technique.[7]

The Search for Some Underlying Structure

In this section, three analytical techniques will be discussed that can be used when researchers are searching for some underlying structure in the data. These techniques include *principal components* (useful when asking if a set of variables can be reduced to a smaller number of "super variables"), *factor analysis* (useful when responses to a set of questions are hypothesized to be driven by just a few underlying factors), and *structural equation modeling* (useful when estimating the set of relationships between one or more independent variables and one or more dependent variables). Although all three techniques involve creating linear combinations of either observed or latent variables, principal components operates exclusively at the empirical level, while the other two are largely theoretical approaches. However, because principal components and factor analysis share many of the same goals and methods, they will be discussed in conjunction with each other, while structural equation modeling will be discussed separately at the end of this section.

Although principal components operates at the empirical level and factor analysis is more of a theoretical construct, both techniques can be applied to a single set of variables when researchers are interested in discovering which variables in the set form coherent subsets that are relatively independent of one another. After these subsets have been identified, the variables within each subset can them be combined into *components* or *factors*, thought to reflect the underlying processes or structure that caused these variables to be correlated with each other. The reason that one technique is said to be empirical while the other largely theoretical, is that factor analysis is typically used when developing and assessing theories, while principal components is traditionally used when trying to reduce a number of correlated variables to a few super variables that preserves most of the variation inherent in the original set of variables.

To illustrate how these techniques might be used, consider the following example from Tabachnick and Fidell (2001) that looks at the characteristics of graduate students. In this example, the researcher collects data from graduate students on their personality characteristics, motivation, intellectual ability, scholastic history, family history, and health and physical characteristics, and then examines the correlations among all of the variables. The patterns suggested by these correlations can then be used to reflect any underlying processes affecting the behavior of graduate students. For instance, several variables from the personality characteristics may be combined with some variables from the motivation and scholastic history measures to form a factor measuring the degree to which a person prefers to work independently – an independence factor. Perhaps several variables from the intellectual ability measures combine with some others from scholastic history to suggest an intelligence factor. In any event, these factors can then be used to develop a theory regarding the behavior of graduate students, which is, of course, an example of factor analysis. However, if the goal of the analysis is to reduce all of these measures into a manageable few that preserve most of the variance in the original measures, then these components – independence and intelligence – can be empirically constructed and then used in any subsequent regression analysis, which is an example of principal components.

So taken together, the goals of principal components and factor analysis can be thought of as summarizing patterns of correlation among observed variables; reducing a large number of variables into a smaller number of components or factors; providing an operational definition for an underlying process by using observed variables; and testing theories about the underlying structure of the data. However, despite the multiple ways in which these techniques can be used, there are a number of problems that researchers must grapple with. Perhaps the most pervasive problem is in trying to identify or give meaning to the observed factors, since they don't always shake out as cleanly as in the previous example. Other problems include the lack of a criterion variable to test the solution[8] and a host of issues involving the infinite number of mathematical rotations that still preserve the same amount of variance but involve different factor loadings. As such, quantitative researchers should be explicit in how they intend on using techniques like these in their proposal, since employing a technique like factor analysis after the fact suggests that they might be trying to "save" poorly constructed or designed research.

Unlike principal components and factor analysis, structural equation modeling is actually a collection of techniques that includes multiple regression analysis, logistical regression analysis, and factor analysis. These techniques can be used with multiple independent and dependent variables, latent or observable variables, as well as discrete or continuous variables. The generality of this collection of techniques has made structural equation modeling[9] probably the hottest analytical technique of the past decade, resulting in some incredibly insightful pieces of quantitative research as well as some of the most poorly designed studies imaginable. Although the learning curve for these techniques is fairly steep since researchers have to first master all of its component parts, the effort is well worth it for those interested in uncovering the underlying structure for a particular process.

However, before actually attempting to use structural equation modeling, it is absolutely crucial to realize that these techniques are confirmatory, rather than exploratory, and should be used to test the appropriateness of various theories. So before ever estimating a model, researchers need to pay careful attention to what theory has to say about the relevancy of particular variables as well as the hypothesized relationships among the variables included in the model. This point is absolutely crucial, because in some circles, structural equation modeling has already developed a bad reputation as a result of researchers spending too much time using these techniques in an exploratory manner, typically testing a variety of different modeling specifications. So researchers interested in using these techniques are urged to focus on testing the appropriateness of a particular theory, rather than on searching for the model that appears to fit the data the best.

In actually using the techniques of structural equation modeling, researchers are also urged to provide a visual representation, or path diagram, for the structure of their proposed model. These diagrams represent the first step in the modeling process since they suggest what sorts of relationships are hypothesized to exist between all of the variables, as well as offering a visual roadmap for the work that follows. After the model has been specified, researchers can then move on to estimation and evaluation, perhaps followed by another path diagram and more estimation and evaluation. Throughout this process, researchers must remember that their ultimate goal is to test the appropriateness of a particular theory and the underlying structure that supports it. If this simple rule is followed and if the researcher has truly mastered the techniques required for structural equation modeling, then the resulting product has the potential to inform not only theorists, but policymakers as well.

The Time Course of Events

For research questions dealing with the time course of events, there are basically two types of admissible analytical strategies. The first of these is called *survival/failure analysis*, and is actually a collection of techniques that can be used to estimate the time that it takes for something to happen. Examples might include the number of semesters it takes for a student to graduate, or the number of years it takes for a student to pay off their student loans. The second type of analytical technique used to study the time course of events is called *time-series analysis*, and is typically used to forecast future events or to evaluate the effect of a time-varying intervention. For example, a time-series model might be built to forecast tuition and fees at a particular private college or to measure the effectiveness of a scholarship dollars on the persistence of a particular group of students. Since both of these analytical strategies are becoming more widely used in a number of fields, including educational research, this section will attempt to provide some analytical context for researchers using this set of techniques in their empirical work.

Within the set of analytical techniques known as survival/failure analysis, there are two very different sorts of procedures – one that uses life or actuarial tables to describe the survival (or failure) time for whatever is being studied, and one that looks at survival time as a function of other variables. While the former procedure uses life tables and hazard functions to calculate the percentage of a given population that survives to some interval in time, the later procedure is more like logistic regression in that survival time can be predicted from a set of variables and uses a log-linear model, which is more forgiving in terms of assumptions about censoring.[10] And just as with logistic regression, the effect size associated with a particular regression coefficient can be interpreted as the natural log of the odds ratio, so that the same sort of calculations described in the logistic regression analysis section are required to calculate effect size. As such, researchers are urged to proceed with caution since there are many well-described, potential analytical pitfalls.

Unlike survival analysis, time-series analysis is typically used when researchers are interested in uncovering patterns in the series, studying the effects of a particular time-dependent intervention, or predicting values for the dependent variable as a function of the model's independent variables. As with standard regression models, researchers are interested in decomposing variation in the dependent variable; in this case the variation might be explained by trends over time, the lingering effects of earlier values of the dependent variable, the lingering effects of earlier shocks, or random variation. Models used for these purposes are often described as an ARIMA (p, d, q), where ARIMA stands for *a*uto-*r*egressive, *i*ntegrated, *m*oving *a*verage, and the *p* represents the lingering effect of preceding values, the *d* represents trends in the data, and the *q* represents the lingering effects of preceding shocks. Since any combination of these effects can be present in a particular model, the researcher first needs to identify which of these patterns may be present so that the model can be properly estimated, and then the residuals or errors examined for randomness. This last step, known as the diagnostic step, is important because if all of the patterns have been successfully removed from the data, then the remaining error terms should be completely random, and if not, then the model may need to be re-estimated. However, in many ways, working with time-series models requires the strongest intuition of any of the techniques discussed in this paper, since decomposing a time-series into time-varying patterns requires years of practice. As time-series guru Clive Granger wrote in his classic textbook, *Forecasting Economic Time Series*:

> "There can be little doubt that the most difficult step in the model building cycle is identification. This is so since, although a number of general principles can be laid down, there exists no surefire deterministic approach to the problem. Rather, it is necessary to exert a degree of judgment, the facility for which is greatly improved by experience. Indeed it has been said that identification is a technique that should not be attempted for the first time" (Granger and Newbold, 1977)

As such, aspiring researchers interested in employing time-series analysis in their work with the Foundation would be well advised to not make this their first foray into this set of techniques, given the idiosyncratic nature of the work. However, the important research questions that can be addressed with time-series analysis makes the set of techniques a valuable addition to any methodologists toolbox, albeit one that takes a substantial amount of time to master.

Research Question #3: What types of data problems can researchers expect to encounter and what sorts of solutions exist?

Now that the five types of research questions and the accompanying techniques have been presented, the focus in this section will be on handling missing data, dealing with unusual observations, and what to do when two or more variables in a database are highly correlated. However, since the first two problems are common to all five types of research questions, more attention will be paid to these particular problems, beginning with several solutions to the missing data problem and their respective statistical properties.

Without a doubt, the problem of *missing data* is the most pervasive problem in all of quantitative research, and at one time or another, all researchers have been forced to deal with the issue. The problem can occur either systematically or randomly, and of course the former is much worse than the later, since if the same survey question regarding household income is continually left blank by those with high levels of income, this may introduce a bias regarding the effect of income in your model, but if survey questions appear to have been left blank randomly, the only modeling cost is a loss of efficiency or precision.[11] In dealing with these sorts of problems for the independent variables in a model, there are really three very different sorts of solutions, including dropping the missing observations, using the sample mean in place of the missing observations, and finally, using sophisticated techniques that involve using other models to forecast the missing observations. Of course, if the missing observations are for the dependent variable in your model, the researcher has no choice but to drop the missing observations from your analysis.

Of the three different correction procedures, the easiest and most convenient method is to simply *drop any observations that you may be missing data for*. Under this correction procedure, if you were collecting thirty bits of data from each individual and if you were missing *any* one of them, then the individual would be completely dropped from the analysis. As mentioned in the preceding paragraph, if the data were missing randomly then the analysis would not be as precise as if you had all of the data; however, if the data were missing systematically, then the analysis would be biased. In terms of trading off between biasness and loss of precision, so long as you still have enough observations to provide relatively precise estimates, then a small loss in precision is much better than introducing a systematic bias into your analysis.

The second type of correction procedure is called the *zero-order correction procedure* and simply involves substituting the sample mean in for the missing observations. From an operational standpoint, this means calculating the sample mean for the variable with the missing observations, and then using it in place of the missing observations. Despite the seeming non-sophistication of this approach, this correction procedure has desirable statistical properties in that it produces unbiased parameter estimates when using regression analysis with only a small loss in precision.[12] Unfortunately, the more variation in the data, the greater the loss in precision, so when researchers are facing cross-sectional databases with lots of variation, this technique may not be as favorable as the first-order correction procedure, discussed below.

The final procedure, known as the *first-order correction*, involves actually estimating other models to predict or forecast the missing observations. Under this procedure, the researcher looks for independent variables that are correlated with the variable that has missing observations but relatively uncorrelated with each other, and then a regression is run on all of the complete cases with the variable with the missing observations serving as the dependent variable, and the correlated variables serving as the independent variables. Sometimes, the predicted values from the first round are inserted in place of the missing observations and then all cases are used in a second regression. This process can then be repeated until the predicted values from each subsequent regression converge.

Although this approach may indeed introduce a small amount of bias into the model, the gain in precision can be large, especially if a number of observations are missing and the regression used for forecasting solid in terms of goodness-of-fit.

In addition to the problem of missing observations, another pervasive data problem concerns what to do about unusual observations, often called *outliers*. These values typically occur for one of four general reasons; incorrect data entry; failure to correctly specify missing value codes so that the codes are incorrectly being read as data; a legitimate value for someone that should not have been sampled for the study (like a staff member at a college university that inadvertently fills out a faculty survey and lists their highest degree as associates); or a legitimate value that just seems unusual (like a 12 year-old college student). Although the first three of these cases are easy to deal with, the case of the 12 year-old college student presents real problems since it appears to be a legitimate observation and may have more influence on your results than appears to be warranted.

In fact, the presence of outliers can have fairly extreme consequences for many forms of quantitative analysis, but especially when conducting regression analysis since the typical regression line is fit by minimizing the sum of squared errors, which of course means that when an observation is far away from the other data points, its contribution to the regression line is huge because of the squared distance measure. Given the problems that outliers can cause, it comes as no surprise that several regression diagnostic measures have been invented that can help in the identification of outliers. These diagnostic measures are called *case statistics*, meaning that there will be one value for each of these statistics for each observation in the database.

The first of these case statistics is called *leverage*, and measures how unusual the case is in terms of its values on all of the independent variables. In other words, this measure gets at how far the observed values for each case are from the mean values on the independent variables. The second type of case statistic is called *discrepancy*, and is a measure of the squared distance between the predicted and observed values for the dependent variable. The third and final case statistic is *influence*, which reflects the amount that the regression coefficients would change if the outlier were removed from the database. For large databases where visual inspection of the data is impossible, researchers are urged to calculate all three case statistics, order them from lowest to highest, and then print out the highest values for examination.

In addition to the use of case statistics, examination of frequency distributions and boxplots may also be helpful in identifying the presence of outliers. However, once an outlier is identified, a determination needs to made as to whether the outlier was caused by some sort of error in the data recording process, or whether the value is legitimate and needs to be dealt with. Of course, if the former is true then the observation can simply be corrected or dropped from the analysis, but if the later is true then there are a limited number of possible solutions. The most popular of these is to redefine the sample, if possible. For example, the presence of a 12 year-old college student in your sample of undergraduates might not be a problem if you redefine your sample to include only those undergraduates that are at least 18 years old. However, if redefining the sample is not an option, then the researcher might consider using some non-linear modeling specification so that the regression line "naturally bends" toward the outlier, rather than pulling the entire regression line towards it, or perhaps even using a least absolute deviation estimator instead of the ones produced through minimizing the sum of squared errors. Regardless of the technique used, researchers need to pay careful attention to the presence of outliers since they have the ability to significantly distort the results of any quantitative study, in effect, limiting the power and generalizability of the study.

In addition to the problems that arise from specific cases within the database, problems can also occur between specific independent variables, most notably the problem of *multicollinearity*, defined as a strong correlation between two or more independent variables. To understand why multicollinearity is a problem, recall that in regression analysis the researcher is interested in the unique, individual contribution that each particular independent variable makes to the dependent variable, and this contribution can be perfectly measured when the independent variables are completely uncorrelated, or orthogonal, with each other. In fact, when we interpret regression coefficients we always have to add the caveat, "holding the other variables constant", which of course cannot be done if a particular independent variable is correlated with another. However, since all independent variables in a model are at least somewhat correlated with each other, the problem of multicollinearity is one of degree, and as such there is considerable debate about exactly when it becomes a problem.

Although there may be some debate as to when multicollinearity becomes a problem, fortunately there is no debate over the consequences of multicollinearity. When multicollinearity occurs between two or more variables, all of the estimated coefficients for these variables have larger standard errors than they otherwise would, which reduces the value of their t-stats and makes it less likely that the researcher will find these variables significant in explaining variation in the model's dependent variable. The two most common ways to test for this problem are the simple bivariate correlation coefficient, which was discussed earlier in this paper, and a measure called the *variance inflation factor* (VIF), which provides an estimate of how much smaller the variance of each of the model's estimated coefficients would be if all of the variables were perfectly uncorrelated. Although there are no precise statistical tests associated with either of these measures, the rule of thumb for the VIF is that any value over 10 provides evidence of serious multicollinearity.

When multicollinearity does occur, there are several accepted methods for dealing with the problem. The simplest and most popular method is simply to drop the offending variable or variables from the model, providing of course that the researcher is not introducing any bias into the analysis by dropping a theoretically important variable. Another simple method is simply to collect additional data, since the larger the sample size the smaller the variance surrounding the estimated regression coefficients. When neither of these techniques can be used, researchers have the option of using the method of principal components to combine all of the multicollinear variables into one or two super variables, which by construction will be perfectly uncorrelated with each other.

However, when there is an extremely high amount of correlation within a dataset, the researcher may feel the need to turn towards ridge regression, which decreases the standard errors surrounding the regression coefficients by adding a constant to the variance of each independent variable. Taken together, these four correction techniques offer researchers a choice of analytical procedures that can be used to increase the precision and generality of their estimates.

Research Question #4: After these models have been estimated, how can the results be presented in a clear and insightful manner?

Although this research question may seem like an afterthought, in many respects it is the most important section in the paper since methodologically sound educational research typically has policy implications, and if policy-makers can't understand the research themselves, then its value is significantly diminished. To help reduce this problem, this section will first discuss what types of information need to be presented so that the results of the analysis are complete and transparent. After readers know *what* to present, the final part of this section will concentrate on *how* to present the information so that both practitioners and scholars can easily follow the story line and the results.

The question of exactly what to report from an analysis has for years confounded many quantitative researchers in education. On the one hand, some researchers feel that every single exploratory model must be reported on, while others are content to report only on their final model. Clearly, the solution lies somewhere in the vast middle, but unless the researcher has discussed the inferential robustness of their results, then as Leamer (1983) argues, "all concepts of traditional theory utterly lose their meaning by the time an applied researcher pulls from the bramble of computer output the one thorn of a model he likes best, the one he chooses to portray as a rose…the consuming public is hardly fooled by this chicanery".

Fortunately, the solution to the problem of inferential robustness is rather simple to implement, although for many researchers it requires a major attitudinal shift. Instead of focusing on a single model, researchers need to use theory to specify sets or families of models that incrementally use different measures and different specifications, including linear and nonlinear ones, to estimate a range of effect sizes. In this manner, rather than producing point estimates for each of the variables in a single model, researchers can report a range of effect sizes for each of the categories of variables used in the family of models. For example, if theory suggests that there should be an income effect in a particular model, and if there are three compelling ways of measuring income, then all three should be used separately and a range of effects reported, not just for the income variable but for the other variables in the model as well. In this manner, the robustness of any inference regarding income can be easily seen, and if it turns out that only one measure of income was significant in only one model, then the inference is not robust but fragile, and any generalizations inappropriate. However, if all three measures were significant in almost every model run then the researcher can have confidence in the existence of an income effect, and the size of the effect reported in a range, rather than as a point estimate.

In addition to reporting a range of effect sizes, the notion of inferential robustness also extends to the way in which researchers handle some of the data problems discussed earlier in this paper. For example, if missing data is a serious problem then researchers might want to try several of the correction procedures to see if their results vary significantly across the different procedures. And of course if they do, then their results are clearly a function of the particular correction technique used and as such, are not generalizable. In much the same way, instead of just trying one solution to the outlier or multicollinearity problem, researchers are urged to explore the full range of analytical solutions and then report on the range of results. In this manner, both researchers *and* Foundation officials can see how sensitive the results of a study are to slightly different correction procedures and modeling specifications, making it ultimately easier for decision-makers to base policy on.

If these simple suggestions are followed, methodologists and other researchers will have an easy time making sense of the importance of a particular piece of research. And when researchers combine the notion of inferential robustness with a sound design and correct set of analytical techniques, the resulting research has the potential to influence policymakers at all levels, provided of course that they can *understand* the research that has been presented. To help researchers in this regard, the discussion now turns to the question of *how* to present quantitative research.

Although many educational researchers may feel that they already know how to present their research in a clear and insightful manner, there are some informal, general rules worth reviewing since the stakes are clearly high for both researchers and Foundation officials. Since these rules involve the substance *and* flow of printed information, the discussion will begin with the flow of information and then move to the specifics of such things as using tables and reporting modeling results. However, before turning to the flow of information there is one over-arching rule for all researchers to remember – the importance of writing in a clear and jargon-free manner. Although at one level this may seem obvious, experienced researchers know that it takes real work to express sophisticated analytical concepts in plain, simple English, but for those that do, the dissemination rewards are great.

In addition to writing in a clear and jargon-free manner, the paper must also flow logically and intuitively from beginning to end. To ensure that this happens, researchers are urged to begin with an introductory section that provides a general overview of the issues surrounding the study and demonstrates the significance of the problem. This section should then be followed by a review of the literature that effectively grounds the study in a larger body of research and provides a clear rationale for the study. The researcher can then articulate the study's research questions and choice of analytical techniques, being sure to discuss exactly why a particular analytical technique was chosen. And of course, if the researcher is using data gathered from a sample, all of the particulars of the sample and the instrument need to be discussed as well. In the results section that then follows, researchers can discuss the inferential robustness of their findings, being sure to focus on the generalizability of their results so that Foundation officials and policymakers will understand exactly what message they can take away from the study. For that matter, researchers are urged to address any policy implications from their study directly in the last section of their paper where they offer their conclusions and suggestions for future research.

By following this general format, which of course parallels the format for a quantitative dissertation, researchers can be assured of a logical flow for their work. However, when working with families or sets of models rather than a single model, legitimate questions arise about using tables and reporting modeling results. Fortunately, the answers are fairly simple as well as intuitive and suggest that instead of simply providing point estimates for everything from estimated coefficients to goodness-of-fit measures like R-squared and adjusted R-squared, intervals or ranges should be provided that contain the results of each successive family of models. In fact, if separate tables like this are constructed for every set of models, then several "super" tables can be constructed that effectively roll-up these individual tables into summaries of all the sets of models, so that readers can get a visual sense of how generalizable the results truly are.

Although these tables may take some time to design and construct, researchers are urged *before writing* to carefully think through the exact order that the tables will appear in, so that the results section has been already organized by table before any writing begins. That way, the layout of tables serves as a sort of advance organizer for the results section, forcing the researcher to think carefully about exactly how the notion of inferential robustness will be addressed. And once the tables have been constructed, researchers are reminded that each table needs to be fully discussed in the text, so that readers know how each table is organized and what the main findings from the table are.

So taken together, these suggestions offer the aspiring quantitative researcher some quality control guidelines for presenting their research. By concentrating on the basics – simple, jargon-free writing, an organizational structure that really allows the writing to flow, and an emphasis on the inferential robustness of their results -- aspiring quantitative researchers can produce work that not only speaks to the need of both Foundation officials and policymakers, but to the broader research community as well.

Conclusions

The broad purpose of this paper has been to provide researchers interested in doing work for Lumina Foundation with an overview, or primer, for conducting methodologically sound quantitative research. In fact, the organization of the paper by research question was designed to stress the importance of just four things – the proper alignment of research questions and the databases that support them, the need to optimally match analytical techniques with research questions, the ability to solve data problems in a methodologically rigorous manner, and the skills and strategies needed to write the paper in a way that maximizes value for all stakeholders. The hope is that by paying attention to these four critical areas, prospective Lumina Foundation grantees will be able to produce more inferentially robust and generalizable research, thereby increasing value for Foundation officials, policymakers, and the broader research community.

Although these suggestions for improving the quality of quantitative research have been directed at researchers doing work for Lumina Foundation, they could just as easily have been targeted at the entire educational research community. However, given the increasingly important role that Lumina Foundation has been taking in both state and federal policymaking, the importance of the Foundation producing methodologically sound research cannot be overstated. And hopefully, as more quantitative researchers in education pay attention to these critical methodological issues, when the National Academy of Science decides to once again look at the state of research in education they won't find what they did in 1992 -- "methodologically weak research, trivial studies, an infatuation with jargon, and a tendency towards fads with a consequent fragmentation of effort" (Atkinson and Jackson, 1992).

Endnotes

[1] Although weighting is a powerful and appropriate tool to use when reporting descriptive statistics, care should be taken when weighting the results of regression analysis since without the appropriate correction procedures, the degrees of freedom available for statistical tests will typically be overstated, resulting in an increased probability of rejecting the null hypothesis.

[2] Although the techniques of canonical correlation and multiway frequency analysis can also be used when assessing the degree of relationship among variables, specifically when there are either multiple dependent variables (canonical correlation) or no dependent variables (multiway frequency analysis), these techniques can be exceedingly difficult to apply and interpret, and as such, are not discussed in this overview.

[3] The multivariate correlation coefficient can also be thought of as a bivariate correlation coefficient between the dependent variable and the composite variable created by the independent variables.

[4] As long as the missing variable is correlated with any of the other variables in the model, then the coefficients of those variables are biased and inconsistent.

[5] Since students typically receive federal, state, or institutional grants, this means that the net price that they pay the college or university has been reduced by the amount of the grant aid received. So for a student facing a sticker price of $20,000 that receives a $5,000 institutional grant, their net price is really $15,000.

[6] The three most important assumptions underlying ANOVA are that the populations being compared are normally distributed, that the variances of these populations are equal, and that the observations are statistically independent.

[7] For example, the traditional goodness-of-fit measure, R-squared, has no real counterpart in logistical regression analysis, and measures based on several log-likelihood functions must be used instead.

[8] This problem is somewhat unique in that with regression analysis, the dependent variable acts as a criterion in that the correlation between observed and predicted values can be measured and used as a test of the solution. The same holds true for the techniques of discriminant function analysis, logistic regression, and multivariate analysis of variance in that prediction of group membership serves as a test of the modeling solution. However, with factor analysis and principal components, no test of the solution is possible.

[9] Structural equation modeling is also known in the literature as causal modeling, simultaneous equation modeling, analysis of covariance modeling, path analysis, and confirmatory factor analysis.

[10] Censoring occurs when the event being studied has not occurred by the time data collection ends. Since by definition, some censoring will always occur in this sort of analysis, estimation techniques that can handle this sort of problem are preferred to those that cannot.

[11] This loss occurs because there are fewer observations to estimate the relationship under study, effectively reducing the precision, or efficiency of any inference from the data.

[12] This surprising result follows from the fact that, by construction, the regression always passes through the point of means, so that when the sample mean is used in place of the

missing observations, the slope of the regression line doesn't change and the estimated coefficients remain unbiased.

Bibliography

Atkinson, R.C., and Jackson, G.B. (1992). *Research and Education Reform: Roles for the Office of Educational Research and Improvement*. Washington DC: National Academy of Sciences.

Cohen, J., Cohen, P., West, S.G., and Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Granger, C.W., and Newbold, P. (1977). *Forecasting Economic Time Series*. New York: Academic Press.

Leamer, E.E., "Let's Take the Con Out of Econometrics," *American Economic Review*, March 1983, *73*, 31- 43.

Tabachnick, B.G., and Fidell, L.S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.

Thompson, B. (1999). "Common Methodological Mistakes in Educational Research, Revisited", Invited address presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Fred J. Galloway is currently associate professor in the School of Education at the University of San Diego, where he has also served as associate dean and director of strategic programs. Prior to joining the university faculty, he was project director for the national Direct Student Loan Evaluation project at Macro International, as well as director of federal policy analysis at the American Council on Education, where he represented the interests of the higher education community before the Executive and Legislative branches of the federal government. Dr. Galloway received his bachelor's and master's degrees from the University of California, San Diego, and his doctoral degree in the economics of education from Harvard University.