Using the General Linear Model to Facilitate the Full Integration of Qualitative and Quantitative Analysis: The Potential to Improve Prediction and Theory Building and Testing

Isadore Newman Florida International University
Anthony J. Onwuegbuzie Sam Houston State University
John H. Hitchcock Indiana University

The data analysis phase of mixed methods research studies typically involves a quantitative analysis of quantitative data and a qualitative analysis of qualitative data, called non-crossover analysis. However, such non-crossover mixed analyses have one important flaw: they maintain the traditional dichotomous distinction between quantitative and qualitative analysis. Thus, what are needed are crossover mixed analyses in which the quantitative and qualitative data analyses are fully integrated. It is only by conducting such crossover mixed analyses that synechism—which represents an antidualistic stance wherein dichotomies are seen as being false and binaries are replaced with continua—can occur. Thus, in this article, we outline a crossover mixed analysis approach that involves the full integration of qualitative analyses with regression approaches.

Modeling in general, and statistical modeling in particular, is widely used to help human beings in their daily lives (e.g., predicting weather), and we all use related procedures to a degree to aid our decision making (e.g., Internet searches, financial investing purposes such as retirement). Statistical modeling also is used in specialized fields that have societal benefits (e.g., identifying serial killers, preventing terrorist attacks and other forms of crime). By its very definition, a statistical model is a mathematical representation of observed data using a set of assumptions (that vary by goal); more generally, modeling entails simulating a process, concept or a system with the goal of understanding it (see, for e.g., http://dictionary.reference.com/browse/modeling). In this sense, readers should be comfortable with the idea that modeling is applicable to all forms of research, and is, in essence, a process through which data are summarized and outcomes are predicted. Indeed, theoretical models have been used in qualitative research for some time now (cf. Glaser, 1965; Glaser & Strauss, 1967) and, as noted previously, it is difficult to imagine anyone reading this article who does not use models—either consciously or subconsciously—on a regular basis.

One widely used approach to modeling is the so-called general(ized) linear model (GLM), which is the underpinning for most statistical analyses used in the social sciences (e.g., regression, independent/dependent samples *t* test, analysis of variance [ANOVA], multiple analysis of variance [MANOVA], factor analysis), serving to describe relationships among an observed set of variables (McNeil, Newman, & Fraas, 2012; Schumacker & Lomax, 1996, 2004). In fact, all parametric analyses—univariate and multivariate analyses alike—with the exception of predictive discriminant analyses, are subsumed by the GLM (Cohen, 1968; Henson, 2000; Knapp, 1978; Onwuegbuzie & Daniel, 2003; Roberts & Henson, 2002; Thompson, 1998). The broad goal for any given GLM application is correctly to specify a model so as to most accurately describe a relationship (or set of relationships) among variables, and when a confirmatory research question is at hand, one does so for testing theory. Given the prominence of the GLM, it stands to reason that there is value both in terms of fully describing the context in which modeling is applied (e.g., Newman & Benz, 1998), and being flexible about the sort of variables that might be used in a model (Miles & Huberman, 1994; Onwuegbuzie & Combs, 2010; Onwuegbuzie & Hitchcock, 2015; Sandelowski, Voils, & Knafl, 2009; Tashakkori & Teddlie, 1998).

The practice of collecting and analyzing quantitative and qualitative data within the same study—otherwise known as *mixed methods research* (Ridenour & Newman, 2008) or *mixed research* (Johnson & Onwuegbuzie, 2004)—formally has taken place since 1972, which represents the year when the first article that used the phrase *mixed methods* has been identified (i.e., Parkhurst, Lovell, Sprafka, & Hodgins, 1972). Over the four decades that have ensued, mixed methods research techniques have developed substantially with respect to many phases of the research process. For example, with respect to the research design phase, in the seminal first edition of the *Handbook of Mixed Methods Research* (Tashakkori & Teddlie, 2003) alone, 35 mixed methods research designs were presented, with these and subsequent designs prompting Nastasi, Hitchcock, and Brown (2010) to develop an inclusive framework for conceptualizing and classifying these designs.

However, despite the significant advances in computer-assisted data analysis software, the developments regarding the phase in which quantitative and qualitative data are analyzed—known as the *mixed analysis* phase (Onwuegbuzie & Combs, 2010)—have lagged behind relative to the other phases. In the overwhelming number of mixed methods research studies, mixed methods researchers collect both quantitative data and qualitative data, and during the mixed analysis phase, conduct a quantitative analysis of the quantitative data and a qualitative analysis of qualitative data, which Onwuegbuzie and Combs (2010) labeled as a *non-crossover mixed analysis*. Such non-crossover analysis yields two sets of findings—quantitative findings and qualitative findings, respectively. These quantitative findings and qualitative findings, in turn, generate quantitative inferences and qualitative inferences, respectively, which, subsequently, optimally are combined into either a coherent whole or two distinct sets of coherent wholes—or what Tashakkori and Teddlie (1998) referred to as a *meta-inference*.

Relative to quantitative research studies that involve solely quantitative analysis of the quantitative data and to qualitative research studies that involve solely qualitative analysis of the qualitative data, non-crossover mixed analyses in mixed methods research studies are much broader because they might involve any one of the 58 (current) classes of quantitative data analysis approaches identified by Onwuegbuzie, Leech, and Collins (2011) (cf. Appendix A) combined with any of the 34 qualitative data analysis approaches identified by Onwuegbuzie and Denham (2014) (cf. Table 1 for 23 of these approaches), any of Miles and Huberman's (1994) 19 within-case analysis methods (i.e., partially ordered display [e.g., checklist matrix]; time-ordered display [e.g., critical incident chart]; role-ordered display [e.g., role-ordered matrix], and conceptually ordered display [e.g., variable-by-variable matrix]) and 18 cross-case analysis methods (i.e., partially ordered display [e.g., partially ordered meta-matrix]; case-ordered display [e.g., case-ordered descriptive meta-matrix]; time-ordered display [e.g., event listing], and conceptually ordered display [e.g., effects matrix]), or any of Saldaňa's (2012) 32 coding techniques (e.g., values coding; which involves applying codes consisting of three elements—*value, attitude,* and *belief*—to examine a participant's perspectives or worldviews). We see these as examples of current design options; innovation will likely yield many more approaches. Thus, when conducting non-crossover mixed analyses, mixed methods researchers have thousands of combinations of quantitative and qualitative analyses at their disposal—and this number of combination increases exponentially when classes of nonparametric quantitative data analyses (cf. Hollander, Wolfe, & Chicken, 2014) and classes of Bayesian analyses (cf. O'Hagan & Luce, 2003) are included.

Non-crossover mixed analyses represent an effective way of combining quantitative and qualitative analyses, providing that they appropriately address the underlying research question(s) and, hence, avoid what Newman, Fraas, Newman, and Brown (2002) refer to as a Type VI error (i.e., the analysis not being appropriate for the research question[s]). However, non-crossover mixed analyses maintain the traditional dichotomous distinction between quantitative and qualitative analysis. Although non-crossover analyses will at times be useful, such a dichotomous stance stands in contrast to Newman and Benz's (1998) notion of mixed methods research representing *interactive continuua*. It is also the case that adopting such a stance in mixed methods research studies prevents all members of the mixed methods research team from interacting maximally: specifically, during the mixed analysis process, the members of the mixed methods research team responsible for conducting the quantitative analyses could operate independently of the members of the team responsible for conducting the qualitative analyses. And this independence of researchers within a mixed research team can be very problematic because at worst, it can lead to conflict among team members (Collins, Onwuegbuzie, Johnson, & Frels, 2013), and, at best, does not maximize the *synergy* (Hall & Howard, 2008) that one can obtain from a mixed research team wherein all team members are interacting maximally throughout the mixed research process in general and the mixed analysis phase in particular.

Consequently, what is needed are mixed analyses in which the quantitative and qualitative data analyses are not just combined but integrated. In fact, it is only by conducting mixed analyses that represent a *fully integrated* approach that synechism—which represents an antidualistic stance wherein dichotomies are seen as being false and binaries are replaced with continua (Johnson & Gray, 2010)—can occur. Thus, in this article, we outline a mixed analysis approach that involves the full integration of qualitative analyses with regression approaches (e.g., linear regression, non-linear regression, probit

Table 1. Most Common Qualitative Analyses

Type of Analysis	**Short Description of Analysis**
Constant Comparison Analysis	Systematically reducing data to codes, then developing themes from the codes.
Classical content analysis	Counting the number of codes.
Word count	Counting the total number of words used or the number of times a particular word is used.
Keywords-in-context	Identifying keywords and utilizing the surrounding words to understand the underlying meaning of the keyword.
Domain analysis	Utilizing the relationships between symbols and referents to identify domains.
Taxonomic analysis	Creating a system of classification that inventories the domains into a flowchart or diagram to help the researcher understand the relationships among the domains.
Componential analysis	Using matrices and/or tables to discover the differences among the subcomponents of domains.
Conversation analysis	Utilizing the behavior of speakers to describe people's methods for producing orderly social interaction.
Discourse analysis	Selecting representative or unique segments of language use, such as several lines of an interview transcript, and then examining the selected lines in detail for rhetorical organization, variability, accountability, and positioning.
Secondary data analysis	Analyzing non-naturalistic data or artifacts that were derived from previous studies.
Membership categorization analysis	Utilizing the role that interpretations play in making descriptions and the consequences of selecting a particular category (e.g., baby, sister, brother, mother, father = family).
Semiotics	Using talk and text as systems of signs under the assumption that no meaning can be attached to a single term.
Manifest content analysis	Describing observed (i.e., manifest) aspects of communication via objective, systematic, and empirical means.
Latent content analysis	Uncovering underlying meaning of text.
Qualitative comparative analysis	Systematically analyzing similarities and differences across cases, typically being used as a theory-building approach, allowing the analyst to make connections among previously built categories, as well as to test and to develop the categories further.
Narrative analysis	Considering the potential of stories to give meaning to individual's lives, and treating data as stories, enabling researchers to take account of research participants' own evaluations.
Text mining	Analyzing naturally occurring text in order to discover and capture semantic information.
Micro-interlocutor analysis	Analyzing information stemming from one or more focus groups about which participant(s) responds to each question, the order that each participant responds, the characteristics of the response, the nonverbal communication used, and the like.
Framework analysis	Analyzing inductively to provide systematic and visible stages to the analysis process, allowing for the inclusion of a priori as well as a posteriori concepts, and comprising the following five key stages: (a) familiarizing, (b) identifying a thematic framework, (c) indexing, (d) charting, and (e) mapping and interpreting.

Table 1. Most Common Qualitative Analyses (continued)

Type of Analysis	Short Description of Analysis
Grounded visualization	Examining spatially a combination of referenced data and ethnographic data, in close relationship to each other, and integrating geographic information systems-based cartographic representations with qualitative forms of analysis and evidence, thereby yielding an inductive and critically reflexive scale-sensitive analysis that combines grounded theory and visualization.
Interpretative phenomenological analysis	Analyzing in detail how one or more persons, in a given context, make sense of a given phenomenon—often representing experiences of personal significance (e.g., major life event).
Schema analysis	Searching for cultural schemata (i.e., scripts) in texts, which include identifying semantic relationships between elements of component schemas.
Ethnographic decision models	Building a model of the decision process for a behavior of interest, resulting in a display of data, via decision trees, decision tables, or sets of rules that take the form of *if-then* statements.

Adapted from "Qualitative data analysis: A compendium of techniques and a framework for selection for school psychology research and beyond," by N. L. Leech and A. J. Onwuegbuzie, 2008, *School Psychology Quarterly*, *23,* p. 601. Copyright 2008 by American Psychological Association.

regression, logistic/logit regression, multivariate logistic regression; cf. Appendix A). By *full integration*, we refer to a mixed methods concept whereby researchers do not conceive of distinct quantitative and qualitative phases, nor is there a clear delineation between philosophical and analytic steps. This is consistent with viewing research as a seamless enterprise with qualitative and quantitative research components (cf. Hitchcock & Newman, 2013; Nastasi et al., 2010; Newman & Hitchcock, 2011). A component of full integration can allow for transforming themes into numbers that can be subject to modeling; furthermore, scores can be transformed back into broader themes (see Boyatzis, 1998; Miles & Huberman, 1994; Onwuegbuzie & Combs, 2010; Sandelowski et al., 2009; Tashakkori & Teddlie, 1998 for discussions about quantitizing themes and qualitizing scores).

Conceptual Framework

In stark contrast to non-crossover mixed analyses are what Onwuegbuzie and Combs (2010) term *crossover mixed analysis*. According to these authors, a crossover mixed analysis involves one form of data (e.g., qualitative) being collected or extracted and, then, are subsequently analyzed utilizing techniques traditionally associated with the alternative paradigm (e.g., quantitative). This type of mixed analysis is conducted under the assumption that (a) quantitative and qualitative analyses are not necessarily distinguishable from each other; (b) any differences between quantitative and qualitative analyses do not justify exclusive quantitative analysis of quantitative data and qualitative analysis of qualitative data; and (c) both quantitative analyses and qualitative analyses can address similar research questions (Onwuegbuzie et al., 2011). To this end, crossover mixed analyses necessitate the integration of qualitative- and quantitative-based paradigmatic assumptions and stances. For example, a crossover mixed analysis might involve combining postpositivist assumptions and stances with phenomenological assumptions and stances within the same analytical framework (interestingly, Mayoh and Onwuegbuzie [2015] outlined how postpositivism and descriptive phenomenology share some ontological and axiological parallels). It is this integrative nature of crossover mixed analyses that led Teddlie and Tashakkori (2009) to declare that "We believe that this is one of the more fruitful areas for the further development of MM [mixed methods] analytical strategies" (p. 281). Thus, crossover mixed analyses served as our conceptual framework.

Purpose of Article

In this article, the crossover mixed analyses that we demonstrate involve the extraction of quantitative data from qualitative data that then are subjected to a quantitative analysis, yielding a form of sequential mixed analysis (Onwuegbuzie & Teddlie, 2003). Specifically, we show how qualitative data and its subsequent qualitative analysis can help inform a quantitative analysis, namely, regression. However, one must keep in mind that we are not necessarily mixing philosophies, but we are mixing methods—specifically, we are *integrating* analytic approaches. We identify the set of variables that are derived from the qualitative analysis in the regression equation as QUAL. These are not traditional qualitative variables, but they were derived from the qualitative analysis and then were quantitized by using either nominal scaling (i.e., belonging to an identified theme or not) or by developing a scoring scheme that leads to the theme(s) being transformed to an ordinal scale of measurement. A description of this is presented in the following sections of our article. It is important to remember that, frequently, qualitative analysis ends with theme development. In this article, we are demonstrating how to conduct a richer analysis by taking the theme development to a logical next step—by categorizing the emerging theme(s) and integrating it fully with quantitative variables of interest.

Toward a Framework for Using the General Liner Model to Facilitate the Full Integration of Qualitative and Quantitative Analysis

Typically, in qualitative research studies, the researcher usually ends with the development and discussion of themes. Unfortunately, as noted by Bazeley (2009), some "qualitative researchers rely on the presentation of key themes supported by quotes from participants' text as the primary form of analysis and reporting of their data" (p. 6), leading to "superficial reporting of themes" (p. 13). In contrast, mixed methods researchers attempt to obtain more information from their qualitative data by transforming or converting the qualitative data (e.g., metathemes, themes) into quantitative codes that can be analyzed quantitatively—a process known as *quantitizing* (Miles & Huberman, 1994; Sandelowski et al., 2009; Tashakkori & Teddlie, 1998) or "quantitative translation" (Boyatzis, 1998, p. 129). However, it should be noted that the goal of quantitizing is not to reduce data, but rather to expand data. As such, the enumerated data do not replace the qualitative data, but supplement these data (i.e., yielding richer and thicker data; Geertz, 1973) by allowing the qualitative data to be placed in a more appropriate context.

Depending on the research question(s), themes that emerge from any type of qualitative analysis—for example, any of the 34 qualitative analysis approaches identified by Onwuegbuzie and Denham (2014)—can be coded as categorical variables, and can be used to determine whether or not (i.e., yes = "1"; no = "0") an individual's response fits that category. Kennedy (1992), in the introduction to his textbook on log-linear analysis, indicated that categorical data are qualitative in nature in much the same way as are variables such as gender, race, and ethnicity, which frequently are used by quantitative researchers as variables in regression models. With this in mind, we conceptualize our GLM-based fully integrated mixed analysis as representing the following six stages: (a) Stage 1: collecting the qualitative and quantitative data; (b) Stage 2: coding the qualitative data; (c) Stage 3: identifying themes from the qualitative codes; (d) Stage 4: assessing intra-coder agreement and inter-coder agreement where applicable; (e) Stage 5: quantitizing the themes to yield an inter-respondent matrix; and (f) Stage 6: using ordinary least squares regression approaches via the inter-respondent matrix to estimate relationships to address fully the research questions. Each of these stages is discussed briefly in the following sections.

Stage 1: Collect the Qualitative and Quantitative Data

Once the researcher(s) has identified the research problem, developed the research question(s), and selected the sampling design and research design, the first step is to collect the qualitative and quantitative data. Depending on the research question, the qualitative data can represent one or more of Leech and Onwuegbuzie's (2008) four major sources of qualitative data prevail: talk (e.g., via individual interviews, focus group interviews), observations (e.g., first-hand, second-hand), images (i.e., sill [e.g., drawings, photographs], moving [e.g., videos]), and documents (i.e., printed, digital). It is essential here that data are collected that are compatible with the research question(s) in order to avoid committing what Newman et al. (2002) refer to as a Type VI error. A Type V1 error occurs anytime there is an inconsistency between a design and the intended research question(s). Type V1 error can be avoided by remaining cognizant of the purpose of conducting the research, and it can help to be aware of research typologies (Newman et al., 2002).

Stage 2: Code the Qualitative Data

Once collected, the researcher(s) is ready to conduct the fully integrated mixed analysis. Thus, Stage 2 involves coding the qualitative data generated by each participant (i.e., within-case analysis). This is undertaken by using one or more qualitative analyses (see, for e.g., Miles & Huberman, 1994;

Onwuegbuzie & Denham, 2014; Saldaňa, 2012). We recommend that, whenever possible, a computer-assisted qualitative data analysis software (CAQDAS) program be used (e.g., QDA Miner; Provalis Research, 2014). Such CAQDAS programs greatly facilitate both the coding and ensuing quantitizing of codes and themes.

Stage 3: Identifying Themes from the Qualitative Codes

Once the qualitative data have been coded, themes are generated using the selected qualitative approaches (cf. Onwuegbuzie & Denham, 2014), methods (cf. Miles & Huberman, 1994), and/or techniques (cf. Saldaňa, 2012). For example, if constant comparison analysis (Glaser, 1965) is the approach used, then the researcher(s) would chunk the qualitative data into smaller meaningful components (i.e., chunks); label each chunk with a descriptive title or a code; compare each new chunk of data with previous codes in order to label similar chunks with the same code; and, after all the data have been coded, group the codes by similarity to yield themes.

Stage 4: Assessing Intra-Coder Agreement and Inter-Coder Agreement

As a way of maximizing rigor in mixed research studies (Onwuegbuzie & Corrigan, 2014), whenever possible, both intra-coder agreement (i.e., intra-rater reliability) and inter-coder agreement (i.e., inter-rater reliability) should be ascertained to assess the level of agreement within each coder and among coders, respectively. When more than two raters are involved, a measure of the multiple coder agreement should be used such as multirater Kappa (Siegel & Castellan, 1988) to provide information regarding the degree to which the coders achieved the possible agreement beyond any agreement than could be expected to occur merely by chance. Ideally some form of *sequential* intra-coder agreement and inter-coder agreement should be used in which an intra-coder and inter-coder agreement checks are conducted early in the coding process, followed by negotiations and adjustments among the coders in an attempt to maximize the intra-coder and inter-coder agreement pertaining to the subsequent coding.

Stage 5: Quantitizing the Themes to Yield an Inter-Respondent Matrix

Once the themes have been identified, the next stage is to quantitize these themes (Boyatzis, 1998; Onwuegbuzie & Combs, 2010; Sandelowski et al., 2009). As described by Onwuegbuzie and Teddlie (2003), this quantitization process involves the researcher(s) assigning a score for each participant in the study and for each theme, depending on the extent to which the participant contributes to that theme. The most basic and common way of assigning scores here is by assigning a score of "1" for each participant who contributes to that theme and assigning a score of "0" otherwise. This dichotomization of themes would yield what Onwuegbuzie (2003) refers to as an (*nominal-scaled*) *inter-respondent matrix* (i.e., *participant x theme matrix*) containing a combination of 0s and 1s. Table 2 provides an idea of how a nominal-scaled inter-respondent matrix might look. A more sophisticated way of coding, if this could be undertaken in a reliable manner, would be to assign codes based on the intensity of contribution provided by each participant using, for example, a rating scale (e.g., 4-point rating scale: 0 = no contribution, 1 = some contribution, 2 = much contribution, 3 = maximum contribution) or Likert-format scale (e.g., 4-point Likert-format scale: 1 = strongly disagree that the participant contributed to the theme, 2 = disagree, 3 = agree, 4 = strongly agree). This would yield an *ordinal-scaled inter-respondent matrix*. Or, even more complexly, an *interval-scaled inter-respondent matrix* or *ratio-scaled inter-respondent matrix* can be constructed—with an example of the latter occurring, for example, if contribution to a theme was measured via the number of words or length of time spoken. Table 3 provides an idea of how an interval-scaled inter-respondent matrix might look. Whatever assignment scheme is used, it should be subjected to assessment of intra-coder agreement and inter-coder agreement.

In addition to the quantitized codes, the inter-respondent matrix should include all quantitative data collected for each participant. Depending on the research question, such quantitative data might include demographic variables, cognitive variables, affective variables, and/or personality variables. Depending on how much quantitative data are involved, for both the nominal-scaled inter-respondent matrix and interval-scaled inter-respondent matrix, each person will contribute one or more rows of data that comprise the following:

Qualitative Themes + Quantitized Data + Quantitative Data

Table 2. Example of a Nominal-Scaled Inter-Respondent Matrix Used to Conduct Crossover Mixed Analysis

ID	Theme 1	Theme 2	Theme 3	Theme 4	Theme 5	Theme 6	Theme 7	Theme 8
001	1	0	1	1	0	0		1
002	0	1	1	1	0	1		0
003	0	1	0	1	0	0		1
.	.	.	.	.	.	.		.
.	.	.	.	.	.	.		.
.	.	.	.	.	.	.		.
205	0	0	0	1	1	1		1

Theme 1 = time; Theme 2 = research/statistics knowledge; Theme 3 = interest/relevance; Theme 4 = text coherence; Theme 5 = vocabulary; Theme 6 = prior knowledge; Theme 7 = reader attributes; Theme 8 = volume of reading

Note: If a study participant listed a characteristic that was eventually categorized under a particular theme, then a score of "1" would be given to the theme for the participant's response; a score of "0" would be given otherwise.

Table 3. Example of an Interval-Scaled Inter-Respondent Matrix Used to Conduct Crossover Mixed Analysis

ID	Theme 1	Theme 2	Theme 3	Theme 4	Theme 5	Theme 6	Theme 7	Theme 8
001	1	2	1	4	2	3		1
002	3	1	1	3	2	1		4
003	4	1	2	1	4	1		3
.	.	.	.	.	.	.		.
.	.	.	.	.	.	.		.
.	.	.	.	.	.	.		.
205	1	2	3	4	1	1		4

Theme 1 = time; Theme 2 = research/statistics knowledge; Theme 3 = interest/relevance; Theme 4 = text coherence; Theme 5 = vocabulary; Theme 6 = prior knowledge; Theme 7 = reader attributes; Theme 8 = volume of reading. Note: This matrix reflects use of a 4-point Likert-format scale (e.g., 4-point Likert-format scale: 1 = strongly disagree that the participant contributed to the theme, 2 = disagree, 3 = agree, 4 = strongly agree)

Stage 6: Using Ordinary Least Squares Regression Approaches

Once the nominal-scaled, ordinal-scaled, interval-scaled, or ratio-scaled inter-respondent matrix has been developed, ordinary least squares regression approaches are applied to address fully the research questions. This marks the sixth and final stage of the GLM-based fully integrated mixed analysis process. Specifically, the quantitized themes are converted to a matrix of bivariate associations. Next, the selected ordinary least squares maximum likelihood regression approach is applied to this converted matrix. Prior to applying the ordinary least squares regression approach to the matrix of bivariate associations stemming from the nominal-scaled inter-respondent matrix, the researcher(s) might consider further converting this matrix of bivariate associations to a matrix of tetrachoric correlation coefficients because these coefficients are appropriate to use when one is determining the relationship between two (artificial) dichotomous variables. Tetrachoric correlation coefficients are based on the assumption that each dichotomous variable can be transformed to a normally distributed latent continuous variable with zero mean and unit variance—yielding a normally distributed latent continuous variable (cf. Onwuegbuzie et al., 2007). Note that it is, however, the case that another approach can be applied where the analyst need not conceptualize this sort of transformation. There can be times when a variable is in fact categorical (e.g., whether a person is alive or dead). That is, one might want to conceptualize categorical variables (including dichotomous variables) as not having an underlying continuum. Consider the difference between a biserial versus a *point biserial correlation*, whereas the former deals with continuous variables and the latter handles categorical variables. When using a point biserial, an analyst can use a phi-coefficient as a means for assessing the level of significance between two categorical variables. With that stated each approach is likely to yield similar overall answers (see McNeil et al., 2012).

Whether the ordinary least squares regression approach is applied to the raw matrix of bivariate associations or a transformed matrix of bivariate associations, it can be used to estimate relationships to reflect research questions such as the following:

Question 1. To what extent do the quantitative and qualitative variables account for a significant amount of unique variance in predicting the dependent variable?

Question 2. To what extent do the quantitative variables account for a significant amount of unique variance over and above the variance explained by the qualitative themes?

Question 3. To what extent do the qualitative themes account for a significant amount of unique variance over and above the variance explained by the quantitative variables?

Question 4. To what extent is there an interaction between the quantitative and qualitative variables?

Question 5. To what extent is there a curvilinear relationship between the qualitative and quantitative variables?

Question 6. To what extent is there a curvilinear interaction between the qualitative and quantitative variables?

Question 7. How stable are the weights and how replicable are the results?

In essence then, ordinary least squares regression can be used to test the aforementioned research questions in the following generic model:

$$Y= a_0U+a_1Quan_1 +a_2Quan_2 +\ldots a_nQuan_n +a_{n+1}Category_1 +a_{n+2}Category_2 \ldots a_{n+n}Category_n+E_1 \quad (1)$$

where: Y= a measure of success; U= the unit vector; a_0-a_{n+1} = Partial regression weights
$Quan_1$-$Quan_n$ = Quantitative measures (variables); $Category_1$- $Category_n$= Qualitative Themes
E= Error (residuals)

Heuristic Example

The example that we provide is based on the data collected by Benge, Onwuegbuzie, Mallette, and Burgess (2010). These researchers used mixed methods research techniques to examine 205 doctoral students' levels of reading ability, their perceptions of barriers that prevented them from reading empirical articles, and the relationship between these two sets of constructs. A constant comparison analysis of open-ended responses provided by these doctoral students regarding their perceptions of barriers that prevented them from reading empirical articles revealed the following eight themes: *time* (i.e., all obligations and activities—including family-, employment-, leisure-, and school-related activities—that consume time and limit the amount of time for reading), *research/statistics knowledge* (i.e., being cognizant of and experienced with research skills including methods, designs, library searches; language pertaining to statistical procedures and data analysis), *interest/relevance* (i.e., lack of interest about the topic and perception that the reading is/is not important to the students' field of study), *text coherence* (i.e., the organization of the text, textual supports [i.e., headings, sub-headings, tables], how well the parts of the text [i.e., words, sentences, paragraphs] connect to create a clear representation for the reader), *vocabulary* (i.e., academic expressions, research-related terminology, and terminology specific to particular fields of study), *prior knowledge* (i.e., familiarity with the topic), *reader attributes* (i.e., the students' perception of their abilities to read and to comprehend empirical literature), and *volume of reading* (i.e., the amount of reading required in the students' daily lives). These variables can be deemed as representing situational antecedents, which refer to factors that surround the stimulus—in this case, empirical research articles.

After these emergent themes had been extracted, the researchers quantitized each theme by assigning a score of "1" for a doctoral student, if that doctoral student had listed a characteristic that was eventually unitized under a particular theme; and assigned a score of "0" otherwise (Onwuegbuzie, 2003; Onwuegbuzie & Teddle, 2003)—which yielded an inter-respondent matrix (i.e., *participant x theme matrix*; Onwuegbuzie, 2003) containing a combination of 0s and 1s (cf. Table 2).

Benge et al. (2010) conducted an array of descriptive, exploratory, and inferential analyses using this inter-respondent matrix. However, they did not use any regression techniques. Indeed, they could have used regression techniques to address the seven research questions presented in the previous section, as follows:

Question 1. To what extent do the quantitative and qualitative variables account for a significant amount of unique variance in predicting the dependent variable?

Ordinary least squares regression could have been used to address this research question by using the eight emergent themes as the set of qualitative variables, and the following variables as the set of quantitative variables: age, grade point average, number of credit hours completed, number of college-level statistics courses taken, number of college-level mathematics courses taken, and number of college-level research methodology courses taken. These variables can be deemed as representing environmental variables, which refer to events that occurred in the past. In Benge et al.'s (2010) study, the two dependent variables of interest were reading comprehension (as measured by the Nelson-Denny Reading Test [NDRT]; Brown, Fishco, & Hanna, 1993) and reading vocabulary (NDRT; Brown et al.). Thus, two regression models would be conducted, one model with reading comprehension as the dependent variable, and the other model with reading vocabulary as the dependent variable. Here, a standard multiple regression analysis can be conducted to identify the extent to which the quantitative and qualitative variables account for a significant amount of unique variance in predicting the two dependent variables.

Alternatively, *all possible subsets* (APS) (i.e., *setwise*) multiple regression could be performed wherein all possible models involving some or all of the independent variables are examined. Specifically, in APS regression, separate regressions are computed for all independent variables singly, all possible pairs of independent variables, all possible trios of independent variables, and so forth, until the best subset of independent variables is identified using an a priori criterion such as the maximum proportion of variance explained (R^2), which provides a popularized measure of effect size (Cohen, 1988). Once the best subset of independent variables has been identified, the researchers then would determine the number of quantitative variables and qualitative variables, and examine the overall contribution of the quantitative variables and the overall contribution of the qualitative variables to this final model to answer the research question.

Question 2. To what extent do the quantitative variables account for a significant amount of unique variance over and above the variance explained by the qualitative themes?

This research question can be answered using a hierarchical multiple regression for each dependent variable. This regression technique allows researchers to specify a fixed order of entry for variables in order to control for the effects of covariates or to test the contribution of certain predictors (i.e., qualitative variables) independent of the influence of other variables (i.e., quantitative variables). The latter rationale is applicable here.

Question 3. To what extent do the qualitative themes account for a significant amount of unique variance over and above the variance explained by the quantitative variables?

Similarly, this research question can be answered using a hierarchical multiple regression for each dependent variable. In this case, this regression technique would be used to test the contribution of the quantitative variables independent of the influence of the qualitative variables.

Question 4. To what extent is there an interaction between the quantitative and qualitative variables?

An *interaction* occurs when the magnitude of the effect of one type of independent variable (e.g., Quan_1) on a dependent variable (e.g., reading comprehension) varies as a function of a second type of independent variable (e.g., Category_1). The regression equation used to analyze and to interpret this two-way interaction in a two-variable regression model is:

$$Y = a_0U + a_1\text{Quan}_1 + a_2\text{Category}_1 + a_3\,\text{Quan}_1\text{xCategory}_1 + E_1 \tag{2}$$

where the last term ($\text{Quan}_1\text{xCategory}_1$) is simply the product of the quantitative variable and qualitative (i.e., quantitized) variable. Here a_3 can be interpreted as the amount of change in the slope of the regression of Y on Quan_1 when Category_1 changes by one unit. If a (statistically) significant interaction effect emerges, then the researchers would examine the unstandardized regression coefficients pertaining to all three variables (i.e., quantitative variable, qualitative variable, interaction variable) to construct a prediction equation from them, as well as their structured coefficients. This model can be extended to regression models that contain multiple quantitative variable and/or multiple qualitative variables. Note that if one variable is continuous and one is dichotomous, then an alternative model is needed whereby the interaction is not a simple multiplication of two variables. Rather, it becomes necessary to work with an intercept and slope within each level of the categorical variable (e.g., if an analyst is dealing with a possible interaction between the dichotomous variable *high school dropout status* and the continuous variable *IQ*, then it is necessary to examine the intercept and slope for dropouts and then the intercept and slope for graduates to assess whether there is an interaction). When both variables are categorical, it is necessary to examine whether there is a multiplicative effect that is independent of a main effect. This is undertaken by taking into consideration a full-information model that has information for each cell (consider a 2X3 scenario), and then comparing results to a restricted model with less information (e.g., working with a dichotomous variable only). If a restricted model replicates cell membership that is known from the full model, then this indicates that there is no interaction. If a restricted model cannot replicate cell membership, then there is evidence that there is an interaction between the two categorical variables (see McNeil et al., 2012 for details).

Question 5. To what extent is there a curvilinear relationship between the qualitative and quantitative variables?

To address this research question, a non-linear multiple regression analysis can be conducted. This analysis involves examination of the non-linear correlations between one or more qualitative variables (i.e., quantitized themes) and a single continuous quantitative variable (e.g., age) or between one or more qualitative variables (i.e., quantitized themes) and a single dependent variable (e.g., reading comprehension).

Question 6. To what extent is there a curvilinear interaction between the qualitative and quantitative variables?

To address this research question, a non-linear multiple regression analysis again can be conducted. This analysis involves examination of the non-linear correlations between one or more interaction variables (e.g., $\text{Quan}_1\text{xCategory}_1$) and a single continuous dependent variable (e.g., reading comprehension).

Question 7. How stable are the weights and how replicable are the results?

There are a number of ways to address Question 7. These include but are not limited to: (a) applying cross-validation procedures by splitting a sample in half, and performing initial analyses on one group and then applying to the next; (b) obtaining a new sample to determine whether results replicate. In both cases, replication estimates can be obtained by forcing beta-weights to be similar across two samples; if one sample produces an R^2 that is within 10% of the effect size of the original sample, then the result is considered to be stable; (c) It is possible to generate mathematical estimates of replication with a single sample. See Newman and Hitchcock (2011) and Newman, McNeil, and Fraas (2004) for details.

Conclusions

We contend that fully integrated mixed analyses represented the most comprehensive and synergistic way to analyze data in mixed methods research studies. Unfortunately, there is scant guidance on how to conduct fully integrated mixed analyses. Researchers who have been using GLM approaches have been mixing these types of data and asking these types of questions from the earliest inception of the GLM. What is different is the emphasis that more recently has been put on the type of information that can be gained from a qualitative perspective, and its potential usefulness for increased prediction and understanding. This is what the mixed methods paradigm worldview emphasizes. Thus, in this article, we have provided a framework for conducting one class of fully integrated mixed analyses, which we call GLM-based fully integrated mixed analysis. Specifically, we outlined an eight-stage GLM-based fully integrated mixed analysis. We contend that our framework represents a small step in an attempt to help beginning and experienced researchers alike to conduct a fully integrated mixed analysis, thereby yielding more coherent meta-inferences, which, in turn, make it easier to reach *verstehen*.

Our suggested approach allows researchers empirically to estimate the weighting of the qualitative and quantitative variables, thereby eliminating the guesswork of the relative importance of each for a particular model. That is, one can empirically estimate which sets account for the most variance in the dependent variable(s) of interest. In the context of our heuristic example, the GLM-based fully integrated mixed analysis would determine whether environmental variables (i.e., quantitative variables) or situational variables (i.e., qualitative variables), in general, are more important in predicting reading ability (i.e., reading comprehension, reading vocabulary). Our fully integrative approach highlights the idea that there is limited value in concentrating on the demarcation between quantitative and qualitative research (Onwuegbuzie, 2012), and, instead, facilitates the ability to align the research questions of interest with the research design and data analyses (i.e., avoid Type VI error; Newman et al., 2002). In future articles, we intend to extend our ideas to curvilinear scenarios. For now, our approach also demonstrates that through the use of the GLM, one can more effectively interpret how the various aspects of the data interrelate and better inform the researcher. Third, because of its integrative nature, our approach facilitates interpretation, prediction, theory building, and theory testing. Finally, our approach helps quantitative and qualitative researchers realize how interdependent their respective analyses are when the research question dictates a mixed methods research study, thereby motivating them to work together closely during the mixed analysis phase, unlike when non-crossover analyses are conducted.

References

Bazeley, P. (2009). Analysing qualitative data: More than 'identifying themes'. *Malaysian Journal of Qualitative Research, 2,* 6-22.

Benge, C., Onwuegbuzie, A. J., Mallette, M. H., & Burgess, M. L. (2010). Doctoral students' perceptions of barriers to reading empirical literature: A mixed analysis. *International Journal of Doctoral Studies*, *5*, 55-77.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny Reading Test: Manual for scoring and interpretation, Forms G and H*. Itasca, IL: Riverside Publishing.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,* 426-443. doi:10.1037/h0026714

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, K. M. T., Onwuegbuzie, A. J., Johnson, R. B., & Frels, R. K. (2013). Practice note: Using debriefing interviews to promote authenticity and transparency in mixed research. *International Journal of Multiple Research Approaches, 7,* 271-283.

Geertz, C. (1973). Thick description toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures* (pp. 3-30). New York: Basic Books.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems, 12,* 436-445. doi:10.1525/sp.1965.12.4.03a00070

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Hall, B., & Howard, K. (2008). A synergistic approach: Conducting mixed methods research with typological and systemic design considerations. *Journal of Mixed Methods Research, 2,* 248-269. doi:10.1177/1558689808314622

Henson, R. K. (2000). Demystifying parametric analyses: Illustrating canonical correlation as the multivariate general linear model. *Multiple Linear Regression Viewpoints, 26*, 11-19.

Hitchcock, J. H., & Newman, I. (2013). Applying an interactive quantitative-qualitative framework: How identifying common intent can enhance inquiry. *Human Resources Development Review, 12*(1)*,* 36-52. doi:10.1177/1534484312462127

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods* (3rd ed.). New York: John Wiley & Sons.

Johnson, R. B., & Gray, R. (2010). *A history of philosophical and theoretical issues for mixed methods research.* In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 69-94). Thousand Oaks, CA: Sage.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26. doi:10.1177/1558689806298224

Kennedy, J. (1992). *Analyzing qualitative data: Log-linear analysis for behavioral research* (2nd ed.). Westport, CT: Praeger Publishers.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85*, 410-416. doi:10.1037/0033-2909.85.2.410

Leech, N. L., & Onwuegbuzie, A. J. (2008). Qualitative data analysis: A compendium of techniques and a framework for selection for school psychology research and beyond. *School Psychology Quarterly*, *23*, 587-604. doi:10.1037/1045-3830.23.4.587

McNeil, K., Newman, I., & Fraas, J. W. (2012). *Designing general linear models to test research hypotheses*. Lanham, MD: University Press of America.

Mayoh, J., & Onwuegbuzie, A. J. (2015). Towards a conceptualization of mixed methods phenomenological research. *Journal of Mixed Methods Research*, *9,* 91–107. doi:10.1177/1558689813505358

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Nastasi, B. K., Hitchcock, J. H., & Brown, L. M. (2010). An inclusive framework for conceptualizing mixed methods design typologies: Moving toward fully integrated synergistic research models. In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of mixed methods in social and behavioral research* (2nd ed., pp. 305-338). Thousand Oaks, CA: Sage.

Newman, I., & Benz, C. R. (1998). *Qualitative–quantitative research methodology: Exploring the interactive continuum.* Carbondale, IL: Southern Illinois University Press.

Newman, I., Fraas, J., Newman, C., & Brown, R. (2002). Research practices that produce Type VI Errors. *Journal of Research in Education*, *12*, 138-145.

Newman, I., & Hitchcock, J. H. (2011). Underlying agreements between quantitative and qualitative research: The short and tall of it all. *Human Resources Development Review,10*, 381-398. doi:10.1177/1534484311413867

Newman, I., McNeil, K., & Fraas, J. (2004). Two methods for estimating a study's replicability. *Midwestern Educational Researcher*, *17*, 36-40.

O'Hagan, A., & Luce, B.R. (2003). *A primer on Bayesian statistics in health economics and outcomes research*. Bethesda, MD: MEDTAP International.

Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity: International Journal of Methodology*, 37, 393-409. doi:10.1023/A:1027379223537

Onwuegbuzie, A. J. (2012). Introduction: Putting the *mixed* back into quantitative and qualitative research in educational research and beyond: Moving towards the r*adical middle. International Journal of Multiple Research Approaches, 6*, 192-219.

Onwuegbuzie, A. J., & Combs, J. P. (2010). *Emergent data analysis techniques in mixed methods research: A synthesis.* In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (2nd ed., pp. 397-430). Thousand Oaks, CA: Sage.

Onwuegbuzie, A. J., & Corrigan, J. A. (2014). Improving the quality of mixed research reports in the field of human resource development and beyond: A call for rigor as an ethical practice. *Human Resource Development Quarterly*, *25*, 273-299.

Onwuegbuzie, A. J., & Daniel, L. G. (2003, February 12). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education* [On-line], 6(2). Retrieved from http://www.dogpile.com/search/web?fcoid=417&fcop=topnav&fpid=27&q=Typology+of+analytical+and+interpretational+errors+in+quantitative+and+qualitative+educational+research&ql=

Onwuegbuzie, A. J., & Denham, M. (2014). Qualitative data analysis. In R. Warden (Ed.), *Oxford bibliographies*. Oxford, England: Oxford Bibliographies.

Onwuegbuzie, A. J., & Hitchcock, J. H. (2015). Advanced mixed analysis approaches. In S. N. Hesse-Biber and R. B. Johnson (Eds.), *Oxford handbook of multiple and mixed methods research*. New York: Oxford.

Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M. T. (2011). Toward a new era for conducting mixed analyses: The role of quantitative dominant and qualitative dominant crossover mixed analyses. In M. Williams & W. P. Vogt (Eds.), *The Sage handbook of innovation in social research methods* (pp. 353-384). Thousand Oaks, CA: Sage.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.

Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M. T., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal*, *44*, 113-160. doi:10.3102/0002831206298169

Parkhurst, P. E., Lovell, K. L., Sprafka, S. A., & Hodgins, M. (1972). *Evaluation of videodisc modules: A mixed methods approach*. (ERIC Document Reproduction Service No. ED348014)

Provalis Research. (2014). QDA Miner (Version 4.1.12) [Computer software]. Montreal, Quebec, Canada: Author.

Ridenour, C. S., & Newman, I. (2008). *Mixed methods research: Exploring the interactive continuum.* Carbondale, IL: Southern Illinois University Press.

Roberts, J. K., & Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement*, *62*, 241-253. doi:10.1177/0013164402062002003

Saldaňa, J. (2012). *The coding manual for qualitative researchers* (2nd ed.). Thousand Oaks, CA: Sage.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research*, *3*, 208-222. doi:10.1177/1558689809334210

Schumacker, R. E. & Lomax, R. G. (1996). *A beginning guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Siegel, S., & Castellan, J. N. (1988). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Applied Social Research Methods Series (Vol. 46). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences.* Thousand Oaks, CA: Sage.

Thompson, B. (1998, April). *Five methodological errors in educational research: The pantheon of statistical significance and other faux pas*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Send correspondence to: Isadore Newman
Florida International University
Email: newmani@fiu.edu

APPENDIX A

Established classes of quantitative data analysis approaches and descriptions.

Measurement Techniques	
Name of Analytical Technique	**Description**
Classical Test Theory	Analyzes the relationship among observed scores, true scores, and error in an attempt to predict outcomes of psychological and behavioral measurement
Item Response Theory (Latent Trait Theory, Strong True Score Theory, Modern Mental Test Theory)	Analyzes the probabilistic relationship between the response that a person provides (e.g.. examinee) on a quantitative item(s) and item parameters (e.g., item difficulty, item discrimination, guessing parameter) and person parameters/latent traits (e.g., ability, personality trait)
Multilevel Item Response	Estimates latent traits of the respondent at different levels and examines the

Theory	relationships between predictor variables and latent traits at different levels
Exploratory Factor Analysis	Explores the underlying structure of correlations among observed variables in an attempt to reduce dimensionality of data, wherein a small(er) number of factors significantly account for the correlations among the set of measured variables; utilizes estimates of common variance or reliability on the main diagonal of the correlation matrix that is factor analyzed
Principal Component Analysis	Explores the underlying structure of correlations among observed variables in an attempt to reduce dimensionality of data, wherein a small(er) number of factors significantly account for the correlations among the set of measured variables; utilizes the total variance of each variable to assess the shared variation among the variables. That is, it uses "ones" on the diagonal of the correlation matrix that is factor analyzed. Principal component analysis typically is conducted for variable reduction because it can be used to develop scores that are combinations of observed variables, whereas exploratory factor analysis is more appropriate for exploring latent constructs and allows for error in estimation models.
Confirmatory Factor Analysis	Verifies the factor structure of a set of observed variables; it allows testing of the hypothesis that a relationship between observed variables and their underlying latent constructs exists
Multiple Factor Analysis (optimal scaling, dual scaling, homogeneity analysis, scalogram analysis)	Analyzes observations described by two or more sets of variables, and examines the common structures present in some or all of these set
Hierarchical Factor Analysis	Differentiates higher-order factors from a set of correlated lower-order factors
Assessing One Variable/Participant at a Time	
Descriptive Analyses (i.e., measures of central tendency, variation/dispersion, position/relative standing, and distributional shape)	Summarizes and describes a set of data one variable at a time in quantitative terms
Single-Subject Analysis	Analyzes observations from one or more individuals in which each individual serves as her/his own control (i.e., individual participant is the unit of analysis, although a group such as a classroom also can be the analytic unit); note that it is possible to include several variables at once in a design but analyses typically focus on one variable at a time
Assessing Differences through Variance Analysis	
Independent samples *t* test	Examines the difference between the means of two independent groups
Dependent samples *t* test (paired samples *t* test)	Examines the difference between the means of two groups, wherein the scores in one group is paired or dependent on the scores in the other group
Analysis of Variance (ANOVA)	Partitions the observed variance into components based on different sources of variation; one-way ANOVA examines the equality of several independent groups based on one dependent/outcome variable; factorial ANOVA examines the effects of two or more independent/explanatory/predictor variables and their interactions
Analysis of Covariance (ANCOVA)	Examines whether one or more factors (and their interactions) have an effect or are related to the outcome variable after removing the variance associated with which quantitative predictors (covariates)

Multivariate Analysis of Variance (MANOVA)	Examines whether one or more factors have an effect or are related to two or more outcome variables
Multivariate Analysis of Covariance (MANCOVA)	Examines whether one or more factors (and their interactions) have an effect or are related to two or more outcome variables after removing the variance associated with quantitative predictors (covariates)
Hierarchical Linear Modeling (HLM) (multilevel modeling, mixed effects modeling, covariance components modeling, random-coefficient regression modeling)	Analyzes variance in an outcome variable when data are in nested categories (e.g., students in a class, classes within a school, schools in one school district)
Multivariate Hierarchical Linear Modeling	Analyzes variance in multivariate dependent variables when the covariance structure of the independent variables is of interest
Repeated Measures Analysis of Variance (RMANOVA)	Involves an analysis of variance conducted on any design wherein the independent/predictor variable(s) have all been measured on the same participants under multiple conditions
Mixed Analysis of Variance (Mixed ANOVA)	Examines differences between two or more independent groups whereby repeated measures have been taken on all participants such that one factor represents a between-subjects variable and the other factor represents a within-subjects variable. Observations also may be nested by a unit (e.g., person) where units are generally treated as a between-subject variable.
Repeated Measures Analysis of Covariance (RMANCOVA)	Examines whether one or more factors (and their interactions) have an effect or are related to the outcome variables (i.e., repeated measures) after removing the variance associated with quantitative predictors (covariates)
Assessing Group Membership/Relationships	
Cluster Analysis	Assigns a set of observations, usually people, into groups or clusters wherein members of the group are maximally similar
Q Methodology	Involves finding relationships between participants across a sample of variables
Profile Analysis	Classifies empirically individual observations based on common characteristics or attributes measured by an observed variable(s)
Multivariate Profile Analysis	Classifies empirically individual observations based on common characteristics or attributes (i.e., multiple dependent variables) measured by observed variables (i.e., multiple independent variables)
Chi-Square Analysis	Involves any test statistic that has a chi-square distribution but generally analyzes the independence of two categorical variables via a contingency table
Chi-Square Automatic Interaction Detection (CHAID)	Examines the relationships between a categorical dependent measure (dichotomous, polytomous, ordinal) and a large set of selected predictor variables that may interact themselves; it involves a series of chi-square analyses (i.e., iterative, chi-square tests of independence) being conducted between the dependent and predictor variables
Multivariate Chi-Square Automatic Interaction Detection (CHAID)	Examines the relationships between two or more categorical dependent measure (dichotomous, polytomous, ordinal) and a large set of selected predictor variables that may interact themselves; it involves a series of chi-square analyses (i.e., iterative, chi-square tests of independence) being conducted between the multiple dependent and predictor variables
Descriptive Discriminant Analysis	Explains group separation (i.e., categorical dependent/outcome variable) as a function of one or more continuous or binary independent variables

Predictive Discriminant Analysis	Predicts a group membership (i.e., categorical dependent/outcome variable) by one or more continuous or binary independent variables
Assessing Time and/or Space	
Time Series Analysis	Involves analyzing, using frequency-domain methods or time-domain methods, an ordered sequence of observations over time, taking into account the serial dependence of the observations for the purpose of modeling and forecasting.
Survival Analysis	Analyzes time-to-event data (i.e., failure time data)
Geostatistics	Analyzes spatiotemporal (i.e., existing in both space and time) datasets
Panel Data Analysis	Analyzes a particular participant or group of participants within multiple sites, periodically observed over a defined time frame (i.e., longitudinal analysis).
Correspondence Analysis	Converts data organized in a two-way table into graphical displays, with the categories of the two variables serving as points; this graphical display presents the relationship between the two categorical variables
Canonical correspondence analysis (CCA)	Relates specific variables (e.g., types of species) to variables of interest (e.g., types of environments)
Fuzzy correspondence analysis	Similar to Correspondence Analysis, except uses “fuzzy data”—data that are coded with multiple categories instead of the common “0” or “1”
Multiple Correspondence Analysis	Analyzes the pattern of relationships of several categorical dependent variables
Discriminant Correspondence Analysis	Categorizes observations in predefined groups using nominal variables
Proportional Hazard Model	Estimates the effects of different covariates influencing the times-to-failure of a system (i.e., hazard rate)
Explaining or Predicting Relationships Between Variables	
Linear Regression	Examines the linear correlations between one (simple regression) or more (multiple regression) binary or continuous explanatory variables and a single continuous dependent variable
Non-Linear Regression	Examines the non-linear correlations between one or more binary or continuous explanatory variables and a single continuous dependent variable
Probit regression	Examines the non-linear correlations between one or more binary or continuous explanatory variables and a binomial response variable
Regression Discontinuity Analysis	Examines causal effects of interventions, wherein assignment to a treatment condition is determined, at least partly, by the value of an observed covariate that lies on either side of a fixed threshold/cut-score
Logistic Regression (logit regression)	Examines the relationship between one (simple logistic regression model) or more (multiple logistic regression model) binary or continuous explanatory variables and a single categorical dependent variable
Multivariate Logistic Regression	Examines the relationship between one or more explanatory variables and two or more categorical dependent variable(s)
Descriptive Discriminant Analysis	Explains group separation (i.e., categorical dependent/outcome variable) as a function of one or more continuous or binary independent variables
Predictive Discriminant Analysis	Predicts a group membership (i.e., categorical dependent/outcome variable) by one or more continuous or binary independent variables.
Log-Linear Analysis (multi-way frequency analysis)	Determines which of a set of three or more variables (and/or interactions) best explains the observed frequencies with no variable serving as the dependent/outcome variable
Canonical Correlation Analysis	Examines the multivariate relationships between two or more binary or continuous predictor variables and two or more binary or continuous outcome variables

Path Analysis	Describes and quantifies the relationship of a dependent/outcome variable to a set of other variables, with each variable being hypothesized as having a direct effect or indirect effect (via other variables) on the dependent variable
Structural Equation Modeling (causal modeling, covariance structure analysis)	Involves building and testing statistical models; it encompasses aspects of confirmatory factor analysis, path analysis, and regression analysis
Multilevel Structural Equation Modeling	Used when the units of observation form a hierarchy of nested clusters and some variables of interest are measured by a set of items or fallible instruments
Multilevel latent class modeling	Analyzes data with a multilevel structure such that model parameters are allowed to differ across groups, clusters, or level-2 units; the dependent variable is not directly observed but represents a latent variable with two or more observed indicators
Correlation coefficient	Measures the association between two variables
Multidimensional Scaling	Explores similarities or dissimilarities in data; it displays the structure of a set of objects from data that approximate the distances between pairs of the objects
Social Network Analysis	Involves the identification and mapping of relationships and flows among people, groups, institutions, web sites, and other information- and knowledge-producing units of different sizes; it provides both a visual and a mathematical analysis of complex human systems; the unit of analysis is not the individual, but an element consisting of a collection of two or more individuals and the linkages among them
Propensity Score Analysis	Replaces multiple covariates such that just one score is applied as a predictor rather than multiple individual covariates, thereby greatly simplifying the model; balances the treatment and control groups on the covariates when participants are grouped into strata or subclassified based on the propensity score; it adjusts for differences via study design (matching) or during estimation of treatment effect (stratification/regression)

[a] For many of these analyses, nonparametric versions and Bayesian versions exist.

Note. Adapted from "Toward a new era for conducting mixed analyses: The role of quantitative dominant and qualitative dominant crossover mixed analyses," by A. J. Onwuegbuzie, N. L. Leech, and K. M. Y. Collins, 2011, in M. Williams & W. P. Vogt (Eds.), *The Sage handbook of innovation in social research methods*, pp. 354-356. Copyright 2011 by Sage Publications.