

Data Analysis in Mixed Research: A Primer

Anthony J. Onwuegbuzie (Corresponding author) & Julie P. Combs

Department of Educational Leadership and Counseling

Sam Houston State University, USA

Box 2119, Sam Houston State University, Huntsville, Texas 77341-2119

E-mail: tonyonwuegbuzie@aol.com

Received: March 23, 2011 Accepted: April 20, 2011 doi:10.5296/ije.v3i1.618

Abstract

The purpose of this methodological article is to provide a primer for conducting a mixed analysis—the term used for analyzing data in mixed research. Broadly speaking, a mixed analysis involves using quantitative and qualitative data analysis techniques within the same study. In particular, a heuristic example using real data from a published study entitled “Perceptions of Barriers to Reading Empirical Literature: A Mixed Analysis” (Benge, Onwuegbuzie, Burgess, & Mallette, 2010) is used with the aid of screenshots to illustrate how a researcher can conduct a quantitative dominant mixed analysis, wherein the quantitative analysis component is given higher priority and qualitative data and analysis is incorporated to increase understanding of the underlying phenomenon.

Keywords: Mixed research, Mixed methods research, Quantitative research, Qualitative research, Mixed analysis, Analysis screenshots

1. Mixed Research Defined

Mixed Research, or what is referred to as mixed methods research, involves “mix[ing] or combin[ing] quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study” (Johnson & Onwuegbuzie, 2004, p. 17). As noted by Collins, Onwuegbuzie, and Sutton (2006), mixed research studies contain 13 steps—each of which occur at one of the following three phases of the mixed research process: research conceptualization (i.e., determining the mixed goal of the study, formulating the mixed research objective[s], determining the rationale of the study and rationale[s] for mixing quantitative and qualitative approaches, determining purpose of the study and the purpose[s] for mixing quantitative and qualitative approaches, determining the mixed research question[s]), research planning (i.e., selecting the mixed sampling design, selecting the mixed research design), and research implementation (i.e., collecting quantitative and qualitative data, analyzing the quantitative and qualitative data, legitimating the data sets and mixed research findings, interpreting the mixed research findings, writing the mixed research report, reformulating the mixed research question[s]). Of these 13 steps, analyzing data in a mixed research study potentially is the most complex step because the researcher(s) involved has to be adept at analyzing both the quantitative and qualitative data that have been collected, as well as integrating the results that stem from both the quantitative and qualitative analysis “in a coherent and meaningful way that yields strong meta-inferences (i.e., inferences from qualitative and quantitative findings being integrated into either a coherent whole or two distinct sets of coherent wholes; Tashakkori & Teddlie, 1998)” (Onwuegbuzie & Combs, 2010, p. 398). As such, guidelines and exemplars are needed for conducting mixed analyses. Thus, the purpose of this article is to describe and to illustrate data in mixed research.

2. Mixed Analysis Defined

Mixed analysis is the term used for analyzing data in mixed research. Onwuegbuzie and Combs (2010) recently provided an inclusive definition of mixed analysis that incorporates the definition and typologies that have been presented in major methodological works. These works included articles, book chapters, books, and paper presentations across numerous fields and disciplines such as the social and behavioral sciences (including psychology and education), nursing and allied health, business, and linguistics that spanned 21 years. Based on their interpretations of the extant literature, Onwuegbuzie and Combs (2010) identified 13 criteria that represent decisions that mixed researchers make before, during, and/or after the conduct of their mixed analyses:

1. rationale/purpose for conducting the mixed analysis
2. philosophy underpinning the mixed analysis
3. number of data types that will be analyzed
4. number of data analysis types that will be used
5. time sequence of the mixed analysis
6. level of interaction between quantitative and qualitative analyses
7. priority of analytical components
8. number of analytical phases

9. link to other design components
10. phase of the research process when all analysis decisions are made
11. type of generalization
12. analysis orientation
13. cross-over nature of analysis

Using these 13 criteria, Onwuegbuzie and Combs (2010) derived the following inclusive and comprehensive definition of mixed analysis:

Mixed analysis involves the use of both quantitative and qualitative analytical techniques within the same framework, which is guided either a priori, a posteriori, or iteratively (representing analytical decisions that occur both prior to the study and during the study). It might be based on one of the existing mixed methods research paradigms (e.g., pragmatism, transformative-emancipatory) such that it meets one of more of the following rationales/purposes: triangulation, complementarity, development, initiation, and expansion. Mixed analyses involve the analysis of one or both data types (i.e., quantitative data *or* qualitative data; or quantitative data *and* qualitative data), which occur either concurrently (i.e., in no chronological order), or sequentially in two phases (in which the qualitative analysis phase precedes the quantitative analysis phase or vice versa, and findings from the initial analysis phase inform the subsequent phase) or more than two phases (i.e., iteratively). The analysis strands might not interact until the data interpretation stage yielding a basic parallel mixed analysis, although more complex forms of parallel mixed analysis can be used, in which interaction takes place in a limited way before the data interpretation phase. The mixed analysis can be designed based, wherein it is directly linked to the mixed methods design (e.g., sequential mixed analysis techniques used for sequential mixed methods designs). Alternatively, the mixed analysis can be phase based, in which the mixed analysis takes place in one or more phases (e.g., data transformation). In mixed analyses, either the qualitative or quantitative analysis strands might be given priority or approximately equal priority as a result of a priori decisions (i.e., determined at the research conceptualization phase) or decisions that emerge during the course of the study (i.e., a posteriori or iterative decisions). The mixed analysis could represent case-oriented, variable-oriented, and process/experience oriented analyses. The mixed analysis is guided by an attempt to analyze data in a way that yields at least one of five types of generalizations (i.e., external statistical generalizations, internal statistical generalizations, analytical generalizations, case-to-case transfer, naturalistic generalization). At its most integrated form, the mixed analysis might involve some form of cross-over analysis, wherein one or more analysis types associated with one tradition (e.g., qualitative analysis) are used to analyze data associated with a different tradition (e.g., quantitative data). (pp. 425-426)

Of these 13 decision criteria, the following five criteria appear to be most common: (a) rationale/purpose for conducting the mixed analysis, (b) number of data types that will be analyzed, (c) time sequence of the mixed analysis, (d) priority of analytical components, and (e) number of analytical phases. Each of these criteria is described in the subsequent sections.

Rationale/purpose for conducting the mixed analysis

Greene, Caracelli, and Graham (1989) identified five purposes for mixing quantitative and qualitative data: *triangulation* (i.e., quantitative findings are compared to the qualitative results); *complementarity* (i.e., results from one analysis type [e.g., qualitative] are interpreted to enhance, expand, illustrate, or clarify findings derived from the other strand [quantitative]); *development* (i.e., data are collected sequentially and the findings from one analysis type are used to inform data collected and analyzed using the other analysis type); *initiation* (i.e., contradictions or paradoxes that might reframe the research question are identified), and *expansion* (i.e., quantitative and qualitative analyses are used to expand the study's scope and focus).

Number of data types that will be analyzed

Traditionally, as noted by Creswell and Plano Clark (2007), “Data analysis in mixed methods research consists of analyzing the quantitative data using quantitative methods and the qualitative data using qualitative methods” (p. 128). However, mixed analyses also can involve the sequential analysis of one data type—which are referred to as sequential mixed analyses (Tashakkori & Teddlie, 1998), wherein data that are generated from the initial analysis then are converted into the other data type. For example, a researcher could conduct a qualitative analysis of qualitative data followed by a quantitative analysis of the qualitative codes that emerge from the qualitative analysis and that are transformed to quantitative data (e.g., exploratory factor analysis of themes that emerge from a constant comparison analysis of qualitative data; cf. Onwuegbuzie, 2003). Such conversion of qualitative data into numerical codes that can be analyzed quantitatively (i.e., statistically) is known as quantitizing (Miles & Huberman, 1994; Tashakkori & Teddlie, 1998). Alternatively, a researcher could conduct a quantitative analysis of quantitative data followed by a qualitative analysis of the quantitative data that emerge from the quantitative analysis and that are transformed to qualitative data (e.g., narrative profile formation of a set of test scores or subscale scores representing the affective domain). Such conversion of quantitative data into narrative data that can be analyzed qualitatively is known as qualitizing (Tashakkori & Teddlie, 1998).

Time sequence of the mixed analysis

Time sequence refers to whether the quantitative and qualitative analysis components occur in a chronological order (Creswell & Plano Clark, 2007). Specifically, the qualitative and quantitative analyses can be conducted in chronological order, or sequentially (i.e., sequential mixed analysis), or they can be conducted in no chronological order, or concurrently (i.e., concurrent mixed analysis). When sequential mixed analyses are conducted, either (a) the quantitative analysis component is conducted first, which then drives or informs the subsequent qualitative analysis component (i.e., sequential quantitative-qualitative analysis; Onwuegbuzie & Teddlie, 2003); (b) the qualitative analysis component is conducted first, which then informs the subsequent quantitative analysis component (i.e., sequential qualitative-quantitative analysis; Onwuegbuzie & Teddlie, 2003); or (c) the quantitative and qualitative analyses are conducted sequentially in more than two phases (i.e., iterative sequential mixed analysis; Teddlie & Tashakkori, 2009).

Priority of analytical components

Another important aspect of mixed analyses is the priority or emphasis given to the quantitative analysis component(s) and the qualitative analysis component(s). Either the qualitative and quantitative analysis components can be given approximately equal priority (i.e., equal status) or one analysis component can be given significantly higher priority than the other analysis component (i.e., dominant status). If the quantitative analysis component is given significantly higher priority, then the analysis essentially is a quantitative-dominant mixed analysis, wherein the analyst adopts a postpositivist stance, while believing simultaneously that the inclusion of qualitative data and analysis is likely to increase understanding of the underlying phenomenon (cf. Johnson, Onwuegbuzie, & Turner, 2007). In contrast, if the qualitative analysis component is given significantly higher priority, then the analysis essentially is a qualitative-dominant mixed analysis, whereby the analyst assumes a constructivist-poststructuralist-critical stance with respect to the mixed analysis process, while believing simultaneously that the inclusion of quantitative data and analysis is likely to provide richer data and interpretations (cf. Johnson et al., 2007).

Number of analytical phases

Mixed analyses involve several phases. For example, Greene (2007, p. 155) identified the following four phases of analysis: (a) data transformation, (b) data correlation and comparison, (c) analysis for inquiry conclusions and inferences, and (d) using aspects of the analytic framework of one methodological tradition within the analysis of data from another tradition. Onwuegbuzie and Teddlie (2003) conceptualized a seven-step process for mixed analyses: (a) data reduction (i.e., reducing the dimensionality of the quantitative data and qualitative data), (b) data display (i.e., describing visually the quantitative data and qualitative data), (c) data transformation (i.e., quantizing and/or qualitzing data), (d) data correlation (i.e., correlating quantitative data with quantitized data or correlating quantitative data with qualitized data), (e) data consolidation (i.e., combining both quantitative and qualitative data to create new or consolidated variables or data sets), (f) data comparison (i.e., comparing data from the quantitative and qualitative data sources), and (g) data integration (i.e., integrating both qualitative and quantitative data into a coherent whole).

Heuristic Example

The following mixed research study (Benge, Onwuegbuzie, Burgess, & Mallette, 2010) provides an example of how one can conduct a mixed analysis. This study is relevant to any field because it involves the study of reading ability within the context of doctoral-level research methods courses.

Purpose of the Study

The purpose of Benge et al.'s (2010) study was fourfold: (a) to examine levels of reading ability—as measured by reading comprehension and reading vocabulary—among doctoral students; (b) to identify doctoral students' perceptions of barriers that prevented them from reading empirical articles; (c) to examine the relationship between these perceived barriers and levels of reading vocabulary and reading comprehension; and (d) to determine which

perceived barriers predict the perceived difficulty that doctoral students experience in reading empirical research articles.

Participants were 205 doctoral students enrolled in one of the doctoral-level research design courses at a large research university in the United States. Because all participants contributed to both the qualitative and quantitative phases of the study, and the qualitative and quantitative data were collected concurrently, the mixed sampling design used was a Concurrent Design using Identical Samples (Onwuegbuzie & Collins, 2007). Although in the study the quantitative and qualitative approaches were given approximately equal weight, the researchers placed a greater emphasis on the quantitative analysis phase, yielding a quantitative- dominant mixed analysis. The rationale/purpose for mixing quantitative and qualitative analysis was complementarity and expansion (Greene et al., 1989).

All participants were administered the Nelson-Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993) and the Reading Interest Survey (RIS). The NDRT was used to measure levels of reading vocabulary (80 items; KR-20 = .85) and reading comprehension (38 items; KR-20 = .69). The RIS contains 62 items that are either open-ended (e.g., “What barriers prevent you from reading more empirical research articles?”) or closed-ended (e.g., “Please indicate your perceptions about the levels of ease/difficulty you experience in reading empirical research articles. Please check the option that best applies: 1 = EASY; 2 = SOMEWHAT EASY; 3 = NEITHER EASY NOR DIFFICULT; 4 = SOMEWHAT DIFFICULT; 5 = DIFFICULT”). Figure 1 displays part of these data.

Quantitative Dominant Mixed Analysis: Stage-by-Stage

A sequential mixed analysis (SMA; Onwuegbuzie & Teddlie, 2003; Tashakkori & Teddlie, 1998) was conducted to analyze doctoral students’ test score data and survey responses. This analysis involved six stages.

Stage 1: Quantitative Analysis of Quantitative Data

The first stage involved the use of descriptive statistics (i.e., descriptive stage; data reduction) to compute reading comprehension and reading vocabulary scores and compare them to the normative data. The screenshots for obtaining the descriptive statistics and output are displayed in Figures 2-4. A series of independent samples *t* tests (not shown) revealed that the current sample of doctoral students had statistically significantly higher scores on the reading comprehension ($t = 6.84, p < .0001$; effect size = 0.49) and reading vocabulary ($t = 11.21, p < .0001$; effect size = 0.80) components of the NDRT than did Brown et al.’s (1993) normative sample of 5,000 undergraduate students from 38 institutions. However, disturbingly, approximately 10% of doctoral students attained reading comprehension and reading vocabulary scores that represented the lower percentiles of this normative sample.

Stage 2: Qualitative Analysis of Qualitative Data

In the second stage, the doctoral students’ perceptions of barriers that prevented them from reading empirical articles were subjected to a thematic analysis (i.e., exploratory stage; data reduction) using constant comparison analysis (Glaser & Strauss, 1967). This analysis

revealed the following eight themes that represented students' perceived barriers to reading empirical literature: time, research/statistics knowledge, interest/relevance, text coherence, vocabulary, prior knowledge, reader attributes, and volume of reading.

Stage 3: Quantitative Analysis of Qualitative Data

The themes then were quantitized (i.e., data transformation) such that if a doctoral student listed a characteristic that was eventually unitized under a particular theme, then a score of "1" was assigned to the theme for the student response; otherwise, a score of "0" was assigned. This dichotomization led to the formation of what Onwuegbuzie (2003) called an *inter-respondent matrix* of themes (i.e., *participant x theme matrix*) that consisted only of 0s and 1s. This inter-respondent matrix of 0s and 1s was entered into the SPSS database, alongside the other variables. Figure 5 displays part of these data.

The inter-respondent matrix was used to calculate the frequency (i.e., prevalence rate) of each theme. The steps for conducting the frequency analysis are displayed in Figures 6-8, and the effect sizes pertaining to three of the themes extracted from qualitative data are presented in Figure 9.

Stage 4: Quantitative Analysis of Qualitative Data

The fourth stage of Benge et al.'s (2010) SMA involved a principal component analysis to ascertain the underlying structure of seven of the eight emergent themes (i.e., exploratory stage; data reduction, data display; data consolidation). This analysis was used to determine the number of factors underlying the seven themes. Benge et al. excluded the time theme from the principal component analysis because this theme focused more on life issues (e.g., family, church, coursework) and not reading-related issues (e.g., statistics, vocabulary, familiarity with content). An orthogonal (i.e., varimax) rotation was used. The eigenvalue-greater-than-one rule (i.e., K1; Kaiser, 1958) was used to determine an appropriate number of factors to retain. These factors, or latent constructs, represented *meta-themes* such that each meta-theme contained one or more of the emergent themes (Onwuegbuzie, 2003).

The steps for conducting the principal components analysis on the seven themes are displayed in Figures 10-14. The results of the factor analysis are presented in Figure 15. It can be seen from Figure 15 that the eigenvalue-greater-than-one rule (Kaiser, 1958) suggested a three-factor solution that explained 61.9% of the total variance. The Rotated Component Matrix in Figure 15 indicated that the following themes had pattern/structure coefficients with large effect sizes (i.e., $> .30$) on Factor 1: research/statistics knowledge and interest/relevance; the following themes had pattern/structure coefficients with large effect sizes on Factor 2: prior knowledge, vocabulary, and reader attributes; and the following themes had pattern/structure coefficients with large effect sizes on Factor 3: text coherence and volume of reading. It should be noted that in addition to having a pattern/structure coefficient with a large effect size on Factor 2, vocabulary had a significant but slightly smaller pattern/structure coefficient on Factor 3 (i.e., cross-loading). Further, in addition to having a pattern/structure coefficient with a large effect size on Factor 2, reader attributes had

a significant but smaller pattern/structure coefficient on Factor 3. Benge et al. labeled Factor 1 as *Research Characteristics* (29.03% of the total variance explained; cf. penultimate column in Total Variances Explained in Figure 15), Factor 2 as *Comprehension* (16.65% variance explained), and Factor 3 as *Text Characteristics* (16.22% variance explained).

Stage 5: Quantitative Analysis of Qualitative Data and Quantitative Data

The fifth stage of Benge et al.'s (2010) SMA (i.e., confirmatory analyses; data correlation) involved the correlation between the two reading ability scores (i.e., reading comprehension and reading vocabulary) that were computed in Stage 1 and the seven reading-related themes (i.e., research/statistics knowledge, interest/relevance, text coherence, vocabulary, prior knowledge, reader attributes, and volume of reading) that were extracted in Stage 2 and quantitized in Stage 3 via the inter-respondent matrix. Specifically, these researchers conducted a canonical correlation analysis (Thompson, 1984) to examine the multivariate relationship between the seven reading-related themes and the two reading ability variables.

The canonical correlation analysis (not displayed) revealed a statistically significant and practically significant multivariate relationship ($F[12, 394] = 1.58, p < .05$; Canonical $R_{cI} = .16$) between the seven themes and the two reading ability variables that was mainly characterized by the relationship between vocabulary and reader attributes on one side (i.e., of the set of themes) and reading vocabulary on the other side (i.e., of the set of reading ability variables).

Similarly, Benge et al. conducted a second canonical correlation to examine the multivariate relationship between the three meta-themes extracted in Stage 4 (i.e., Research Characteristics, Comprehension, Text Characteristics) and the two reading ability variables (i.e., confirmatory analyses; data correlation). This canonical correlation analysis (not displayed) indicated a multivariate relationship ($F[6, 400] = 4.37, p < .001$; Canonical $R_{cI} = .34$) between all three meta-themes and reading comprehension. Specifically, these findings suggested that doctoral students with low levels of reading comprehension are more likely to note research characteristics and text characteristics as being barriers to reading empirical articles, but less likely to perceive comprehension as being a barrier.

Stage 6: Quantitative Analysis of Qualitative Data and Quantitative Data

The sixth and final stage of Benge et al.'s (2010) SMA involved a canonical discriminant analysis to determine which of the seven reader-related themes extracted in Stage 2 and quantitized in Stage 3 predicted the perceived difficulty that doctoral students experience in reading empirical research articles (i.e., confirmatory analyses; data correlation). The seven reading-based themes were treated as the predictor set of variables, whereas reading experience (easy vs. neither easy nor difficult vs. difficult) was utilized as the dependent variable. The canonical discriminant analysis (not displayed) revealed that the seven themes statistically significantly and practically significantly predicted the level of perceived difficulty that doctoral students experience in reading empirical research articles ($X^2[12] = 27.41, p = .007$; Canonical $R_{cI} = .36$), with the following variables making the most important contributions to this prediction: reader attributes, interest/relevance, and vocabulary.

A canonical discriminant analysis also was conducted to determine which of the three meta-themes extracted in Stage 4 predicted the perceived difficulty that doctoral students experience in reading empirical research articles (i.e., confirmatory analyses; data correlation). This analysis (not displayed) revealed that the three meta-themes statistically significantly and practically significantly predicted levels of perceived difficulty ($X^2[6] = 18.50$, $p = .005$; Canonical $R_{cI} = .29$), with research characteristics and comprehension making the most important contributions to this prediction.

Based on the array of findings, Bengue et al. (2010) concluded that for doctoral students,

reading ability likely plays an important role in the learning context. Moreover, the negative relationship between levels of reading ability and some of the emergent themes and meta-themes suggests that inadequate reading ability can place a student at risk of not learning the skills necessary to be a consumer of research...by not reading key empirical articles. As such, interventions aimed at improving reading ability among doctoral students likely might help to address their research needs. (p. 48)

Analysis of the Bengue et al.'s (2010) Mixed Analysis

Even though the mixed analysis yielded numerous results, Bengue et al. (2010) could have gone even further with their analysis. For instance, they could have qualitized the data by using the responses to the themes to form narrative profiles. In particular, cluster analysis techniques could have been used to determine the number of clusters (i.e., groups of participants) that underlie their responses. Alternatively, Bengue et al. could have conducted an even stronger quantitative analysis—for example, by using theory to develop a model involving the collected variables (e.g., demographic variables) and then using structural equation modeling techniques to test this model. On the qualitative analysis side, Bengue et al. could have collected and analyzed more qualitative data (e.g., interviews, focus group interviews, observations) to explore further how the seven emergent reading-related themes manifest themselves in the classroom. Nevertheless, the mixed analysis undertaken by Bengue et al. yielded rich findings that provided an impetus for future studies to be conducted in this area.

References

- Bengue, 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.
- 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.
- Collins, K. M. T., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. *Learning Disabilities: A Contemporary Journal*, 4, 67-100.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco: Jossey-Bass.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*, 255-274. doi:10.2307/1163620
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26. doi:10.3102/0013189X033007014, <http://dx.doi.org/10.3102/0013189X033007014>
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*, 112-133. doi:10.1177/1558689806298224, <http://dx.doi.org/10.1177/1558689806298224>
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200. doi:10.1007/BF02289233, <http://dx.doi.org/10.1007/BF02289233>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity: International Journal of Methodology, 37*, 393-409.
- Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12*, 281-316. [Online] Available: <http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf>
- 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., & 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, <http://dx.doi.org/10.3102/0002831206298169>
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches (Applied Social Research Methods Series, Vol. 46)*. 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. (1984). *Canonical correlation analysis: Uses and interpretations*. Newbury Park, CA: Sage.

The screenshot shows the SPSS Data Editor window with the following data table:

Case	Gender	ethnicity	Vocabulary	Comprehension	barriers
1	1	2	2	76	37 time - visual impairment makes additional reading very difficult
2	2	1	1	73	32 Being busy with work.
3	3	1	2	76	36 Time - so much to read and so little time
4	4	2	2	77	37 If the article includes complicated statistical analyses, I find myself rather frustrated
5	5	1	1	75	33 my lack of understanding of the terminology and statistical techniques used in many research articles.
6	6	1	1	75	30 time, sometimes lack of relevancy to work
7	7	2	2	79	36 lack of time, lack of empirical studies in my field
8	8	2	2	78	31 Lack of time. I'm not sure which journals to look at
9	9	1	2	79	37 Time - work, school, family, home ownership
10	10	1	2	74	35 time. I have two small children
11	11	2	2	76	35 Not interesting to me - in some ways the knowledge might not fit my career plans right now.
12	12	1	1	73	32 With a five class schedule, I read what is needed to successfully complete assignments. There is never time left during t...
13	13	1	2	77	26 accessibility and finding articles with the whole text (the study itself)
14	14	1	2	78	30 I still don't have a handle on certain specifics though I understand Chi-sq, anova's, regression, and basic stats
15	15	1	1	76	35 Parental duties, part-time employment obligations, and full time study obligations are three barriers that prevent me from ...
16	16	2	2	77	34 I feel a lack of knowledge in statistical data and methods even though I have had Stats I and 2
17	17	2	2	78	33 vision, balancing life roles and "time", a hang-up with data interpretation
18	18	2	2	71	33 Lack of time to digest them. Lack of time to do much reading at all. Too tired to focus on data and analysis by time I ge...
19	19	2	2	77	36 length of time required to read, full attention required while reading
20	20	2	2	68	34 maybe the topics and long that is use!
21	21	1	2	74	35 poor writing quality of some
22	22	1	1	75	26 English as a second language; vocabulary in English
23	23	2	2	76	35 the results section is usually complicated
24	24	2	2	75	32 technical vocabulary, too much information difficult to determine the bottom line
25	25	1	2	78	37 since my field is social studies, I prefer ethnographic and oral history type of articles. I do read extensively, but not "emp...
26	26	1	2	71	28 balancing time to read research articles, work, homework, and "fun" reading
27	27	2	1	53	31 too long, difficult to understand
28	28	2	1	64	37 If I had a better understanding of psychometrics I would read more empirical research articles
29	29	2	1	70	35 lack of interest in topic
30	30	2	2	74	35 Time - Time - Time - I have to spend to read more articles with required classes and keep my grade, but until then...

Figure 1. Screenshot of data showing a partial view of selected variables: gender, ethnicity, reading comprehension score, reading vocabulary score, perceptions of barriers to reading empirical research articles in the SPSS data editor.

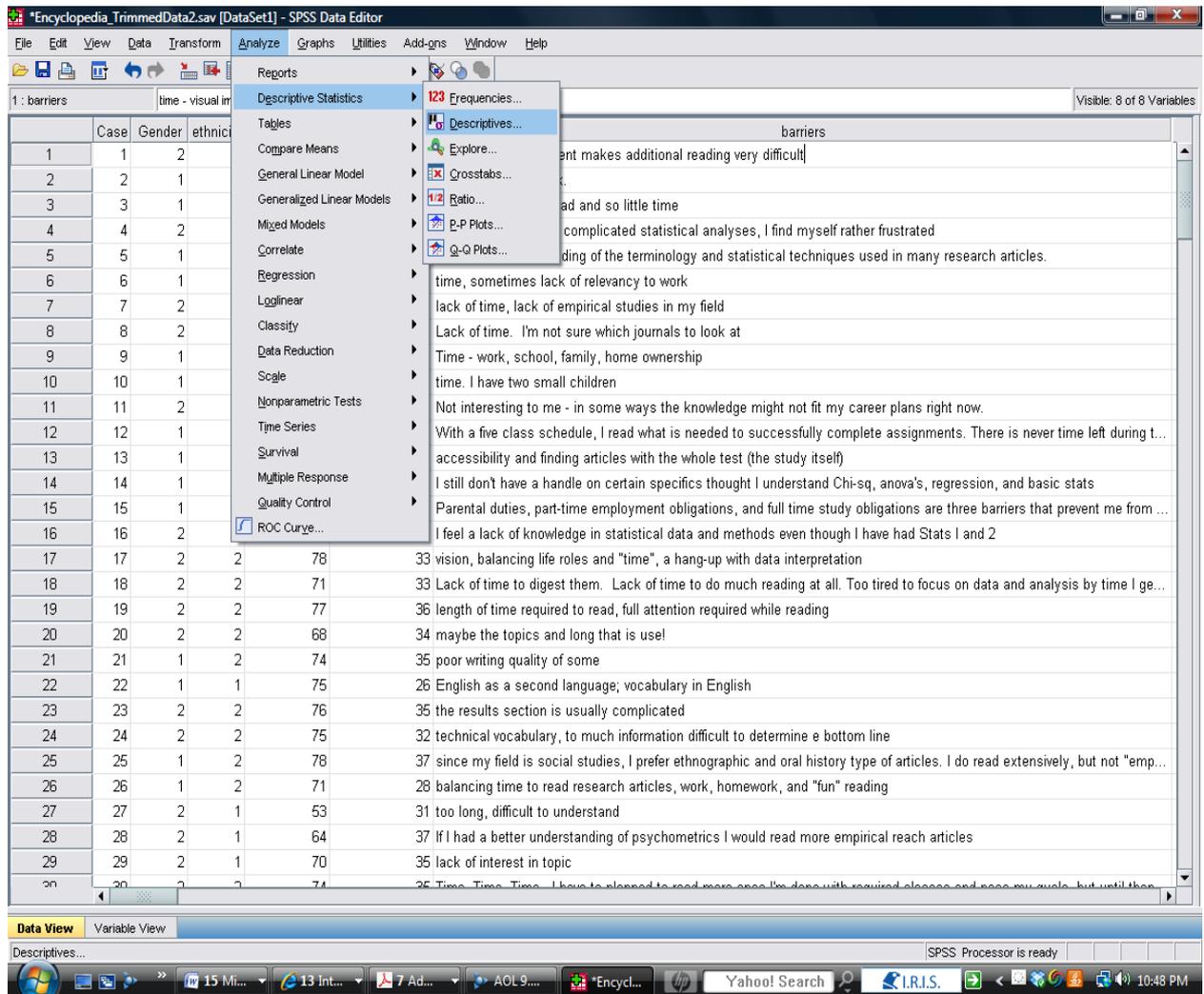
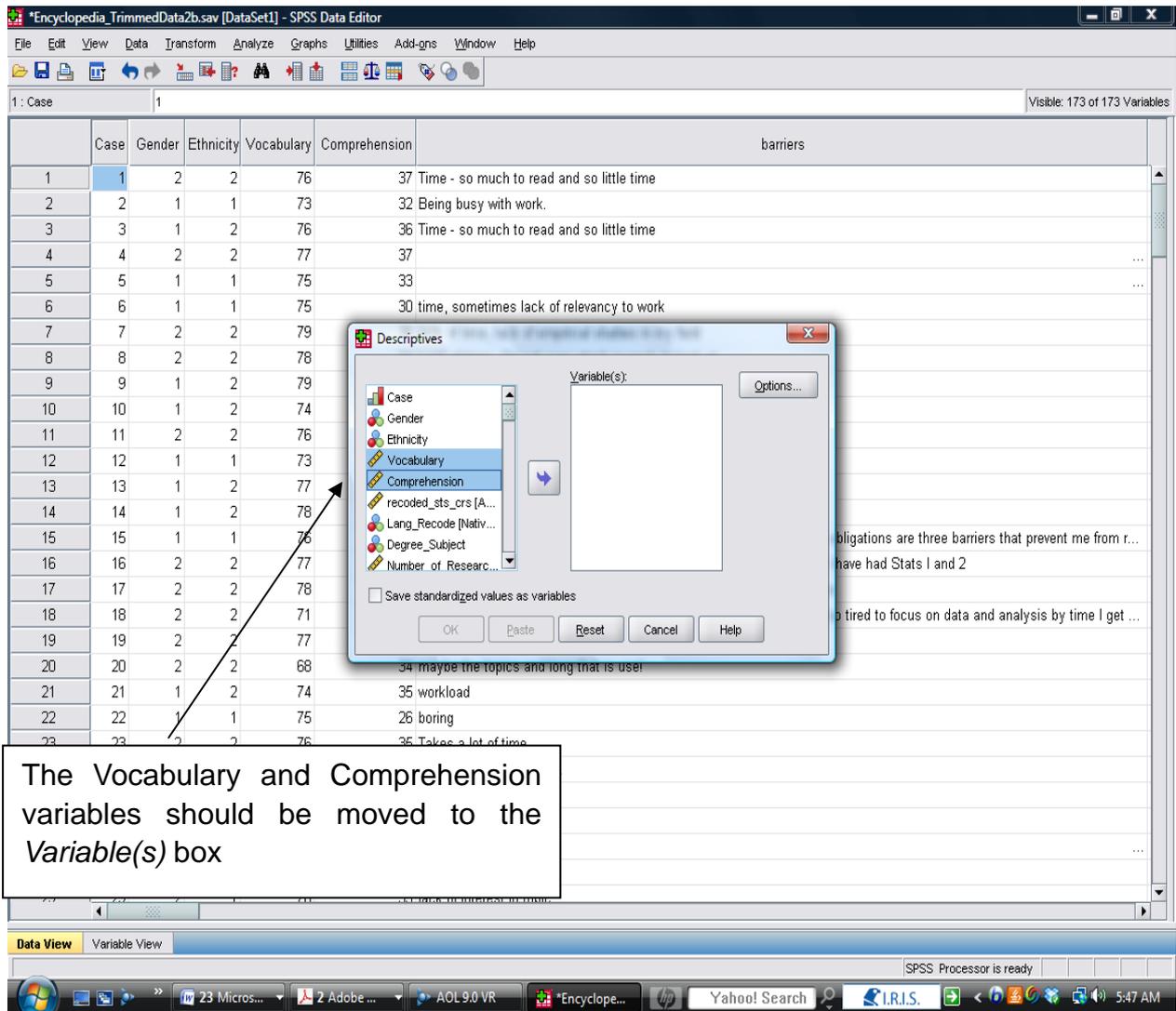
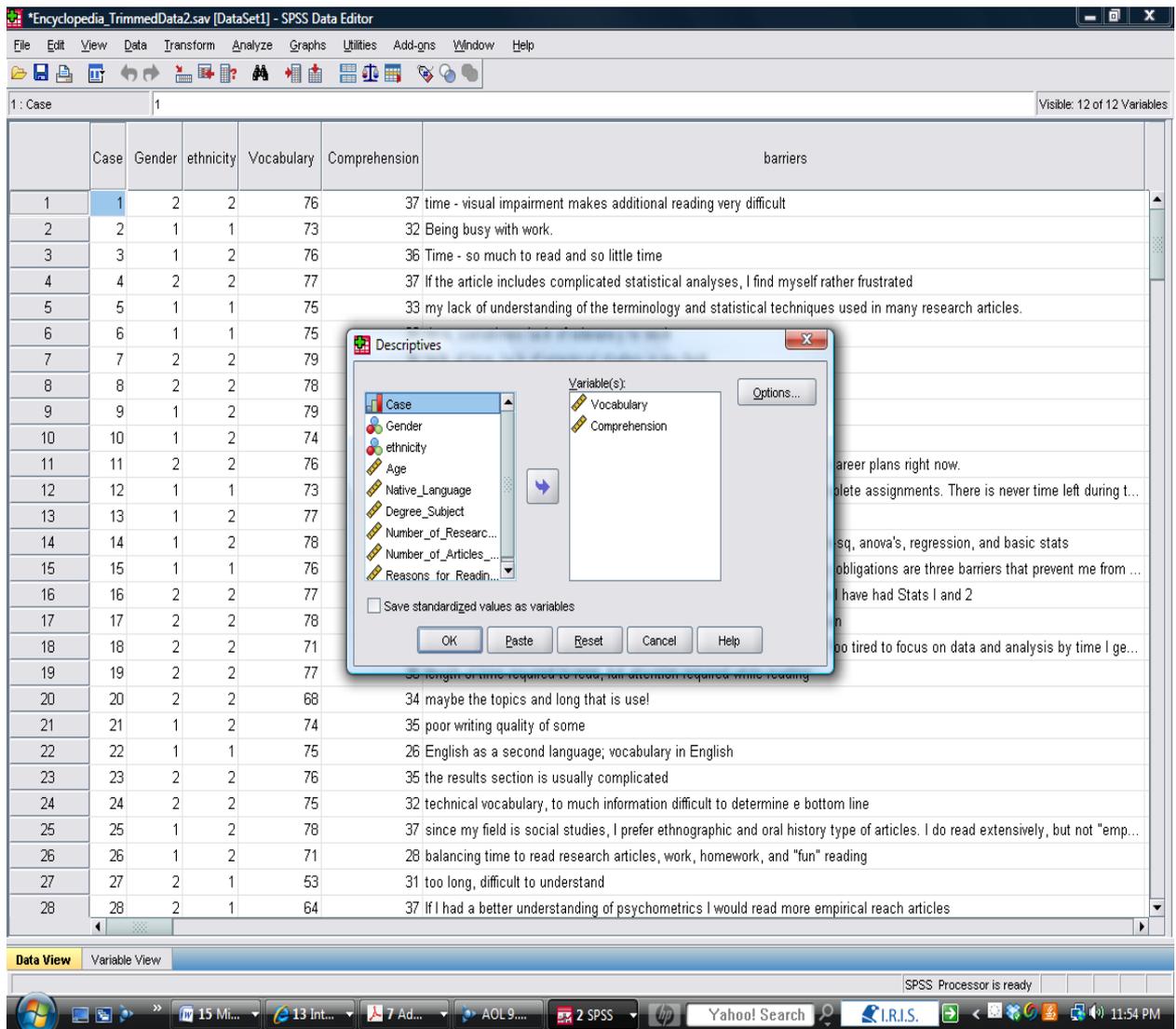


Figure 2. Screenshot showing command path for obtaining descriptive statistics.



The Vocabulary and Comprehension variables should be moved to the Variable(s) box

Figure 3. Screenshot showing Descriptive Box for selecting reading comprehension and reading vocabulary variables for obtaining descriptive statistics



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Vocabulary	205	42	80	73.58	5.948
Comprehension	205	42	76	68.62	6.013
Valid N (listwise)	205				

Figure 4. Screenshot showing Descriptive Box for selecting reading comprehension and reading vocabulary variables for obtaining descriptive statistics, and output of descriptive statistics

	Gender	ethnicity	Vocabulary	Comprehension	Time	Research_Statistics_Knowledge	Interest_Relevance	Text_Coherence	Vocab	Prior_Knowledge	Reader_Attributes	Volume_of_Reading
1	2	2	76	76	1	1	1	0	0	0	1	1
2	1	1	73	73	1	1	1	0	0	0	0	0
3	1	2	76	76	0	0	1	1	0	0	1	1
4	2	2	77	77	0	0	0	0	1	0	0	1
5	1	1	75	75	1	1	0	0	0	0	0	0
6	1	1	75	75	1	1	1	0	0	0	0	0
7	2	2	79	79	0	0	1	0	0	0	0	1
8	2	2	78	78	1	1	1	0	0	0	0	0
9	1	2	79	79	1	1	1	0	0	0	0	0
10	1	2	74	74	0	0	1	0	0	0	0	1
11	2	2	76	76	0	0	1	0	1	0	0	0
12	1	1	73	73	0	0	1	0	1	0	0	1
13	1	2	77	77	1	1	0	0	0	0	0	0
14	1	2	78	78	0	0	1	1	1	0	0	0
15	1	1	76	76	1	1	0	0	0	0	1	0
16	2	2	77	77	1	1	0	0	0	0	0	0
17	2	2	78	78	1	1	1	0	0	0	0	0
18	2	2	71	71	1	1	1	0	0	0	0	0
19	2	2	77	77	1	1	1	0	0	1	0	0
20	2	2	68	68	0	0	0	0	0	0	0	0
21	1	2	74	74	0	0	1	0	0	0	0	0
22	1	1	75	75	.	.	1
23	2	2	76	76	1	1	1	1	0	0	0	1
24	2	2	75	75	1	1	1	0	0	1	0	0
25	1	2	78	78	0	0	1	1	0	0	0	1
26	1	2	71	71	0	0	0	0	0	0	0	1
27	2	1	53	53	0	0	1	0	0	0	0	0
28	2	1	64	64	0	0	0	0	1	0	0	0
29	2	1	70	70	1	1	1	0	1	0	0	1

Figure 5. Screenshot of data showing a partial view of selected demographic variables (i.e., gender, ethnicity), reading vocabulary scores, reading comprehension scores, and the eight themes (i.e., time, research/statistics knowledge, interest/relevance, text coherence, vocabulary, prior knowledge, reader attributes, and volume of reading) in the SPSS data editor.

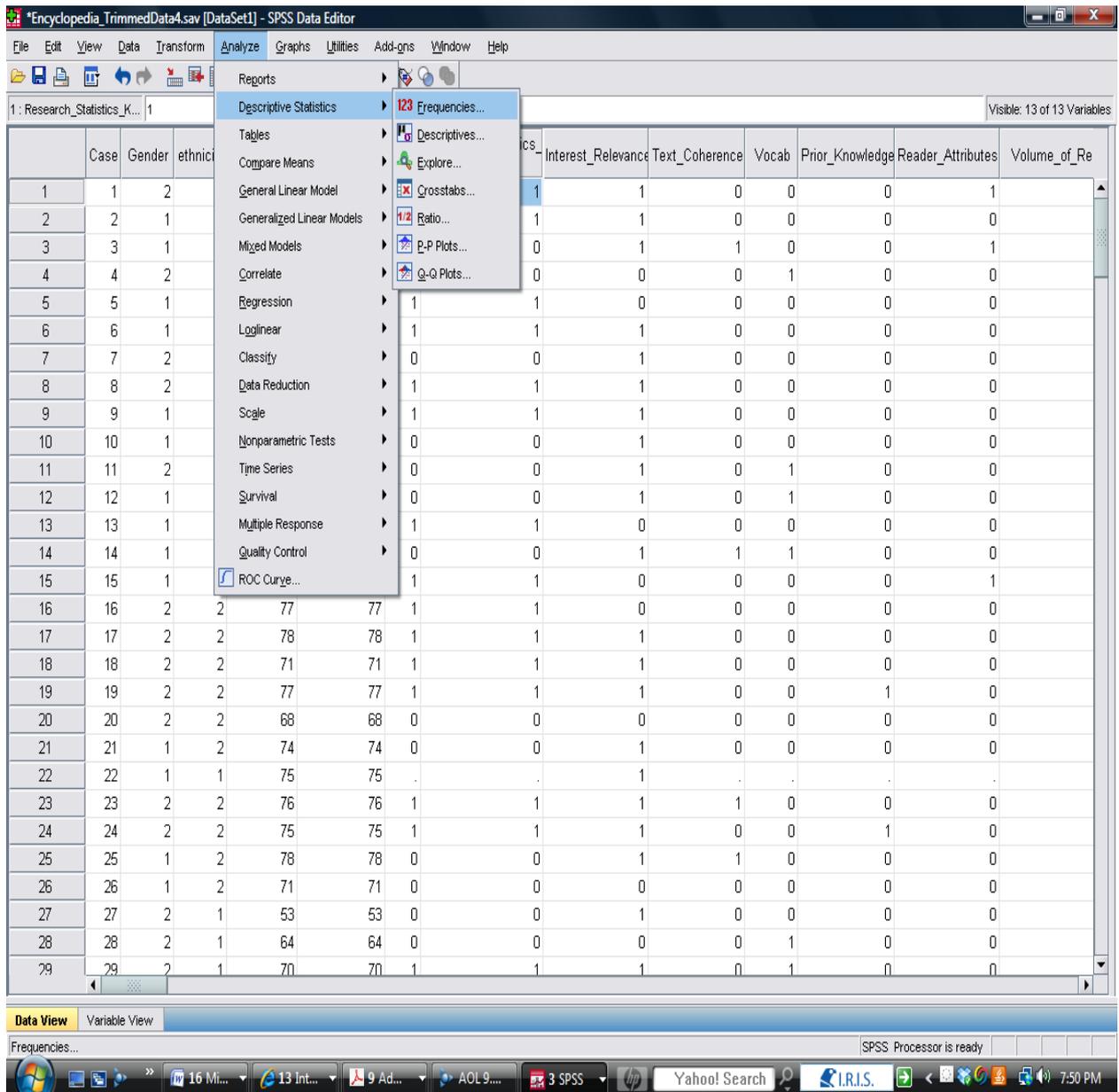
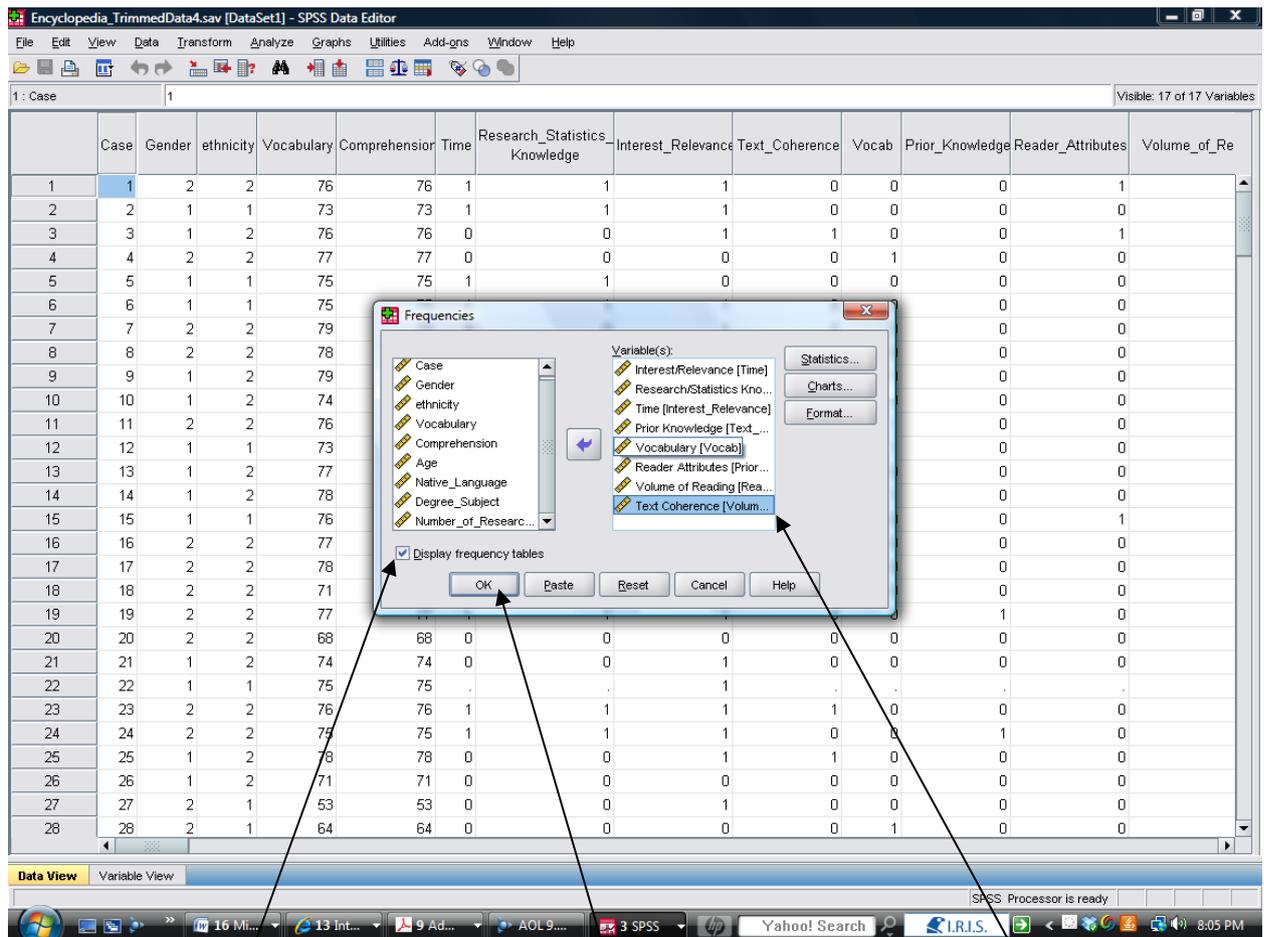


Figure 6. Screenshot showing command path for conducting a frequency analysis

The screenshot shows the SPSS Data Editor window with a data table containing 29 rows and 13 columns. The columns are: Case, Gender, ethnicity, Vocabulary, Comprehension, Time, Research_Statistics_Knowledge, Interest_Relevance, Text_Coherence, Vocab, Prior_Knowledge, Reader_Attributes, and Volume_of_Reading. The 'Frequencies' dialog box is open, showing a list of variables on the left and an empty 'Variable(s):' box on the right. Eight variables are selected in the list: Comprehension, Interest/Relevance [T..., Research/Statistics ..., Time [Interest_Relev..., Prior Knowledge [Te..., Vocabulary [Vocab], Reader Attributes [Pr..., and Volume of Reading [...]. The 'Display frequency tables' checkbox is checked. A callout box at the bottom left of the screenshot contains the text: 'The eight themes should be moved to the Variable(s) box.'

Figure 7. Screenshot showing eight themes that are positioned for a frequency analysis

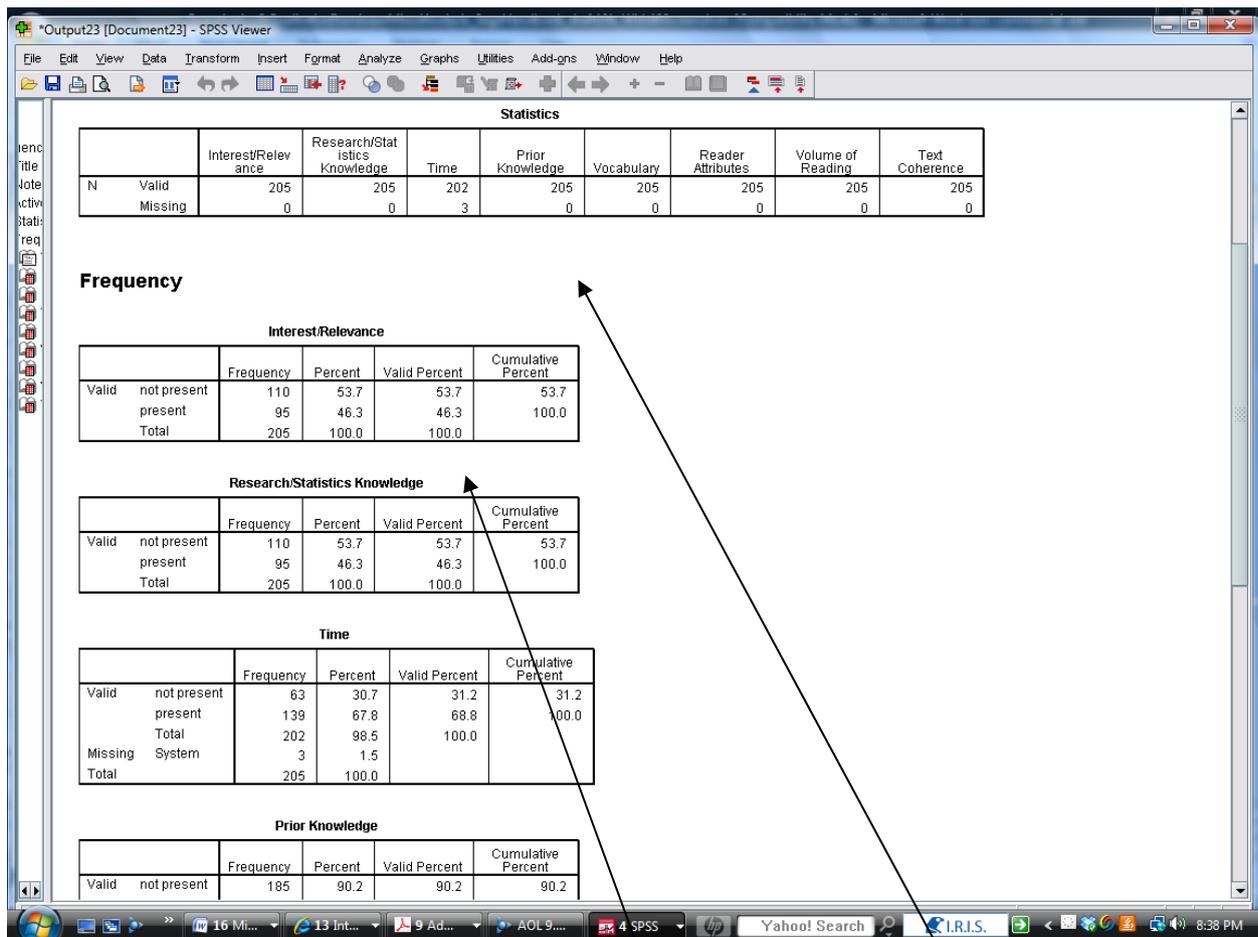


Leaving the *Display frequency tables* option checked provides frequencies of the seven themes.

The eight themes have been moved to the *Variable(s)* box.

Clicking on *OK* computes the frequencies.

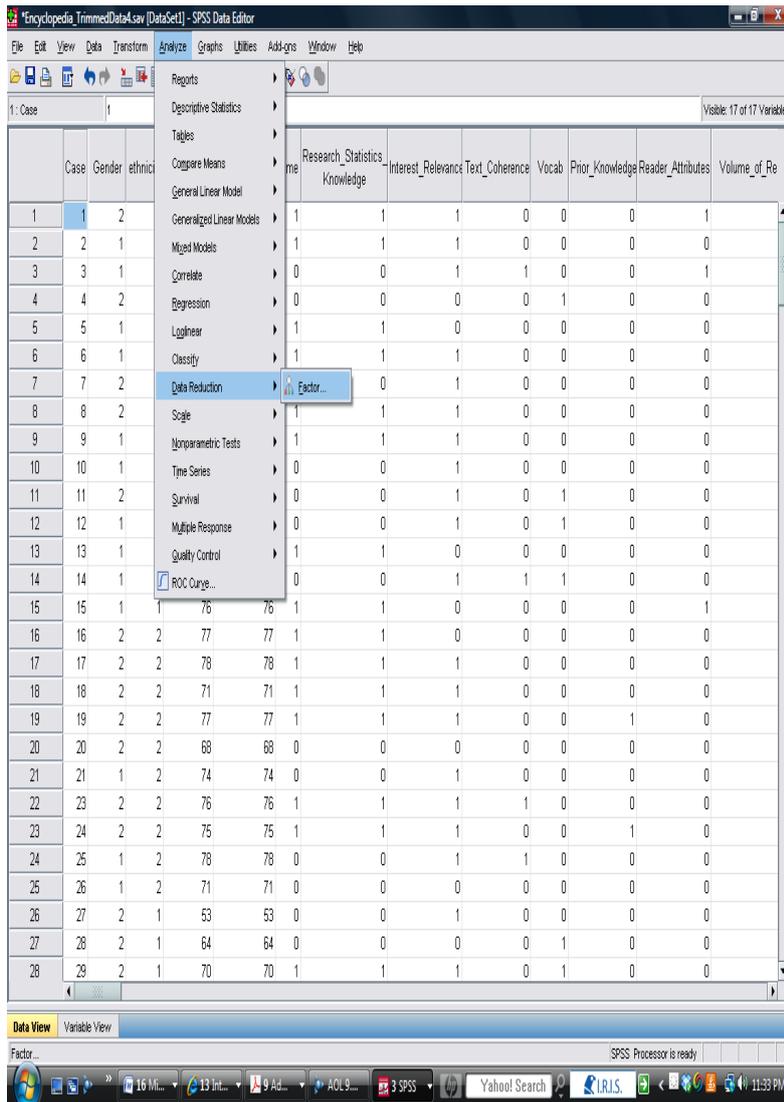
Figure 8. Screenshot showing selection of the themes for the frequency analysis



The “present” percentage indicates the proportion of participants who made one or more statements that were classified under the theme *interest/relevance*. The 46.3% represents the manifest effect size for this theme.

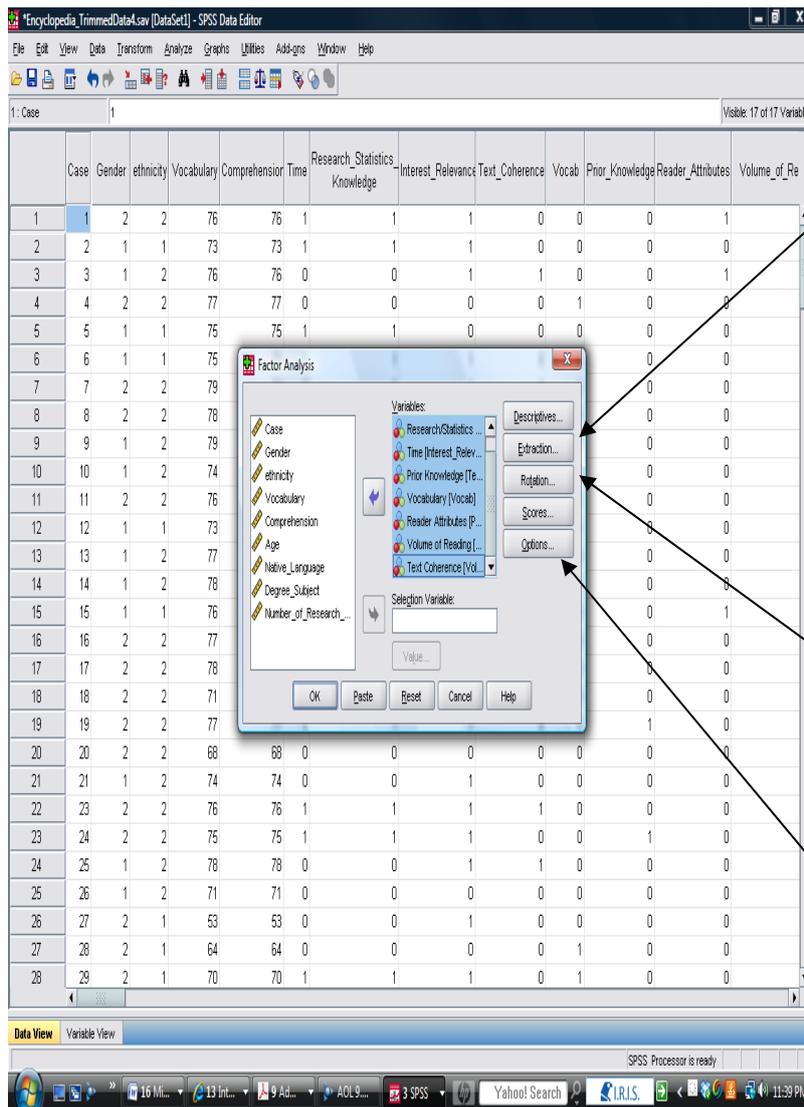
The *Statistics* box shows that only one of the eight themes (i.e., time) had missing data.

Figure 9. Screenshot showing output from the frequencies option yielding effect sizes after quantizing the data (i.e., data transformation)



Ideally, the inter-respondent matrix should be transformed to a matrix of bivariate associations that represented tetrachoric correlation coefficients because the themes had been quantized to dichotomous data (i.e., “0” vs. “1”), and tetrachoric correlation coefficients are appropriate to use when one is determining the relationship between two (artificial) dichotomous variables (cf. Onwuegbuzie, Witcher, Collins, Filer, Wiedmaier, & Moore, 2007).

Figure 10. Screenshot showing the command path for exploratory factor analysis

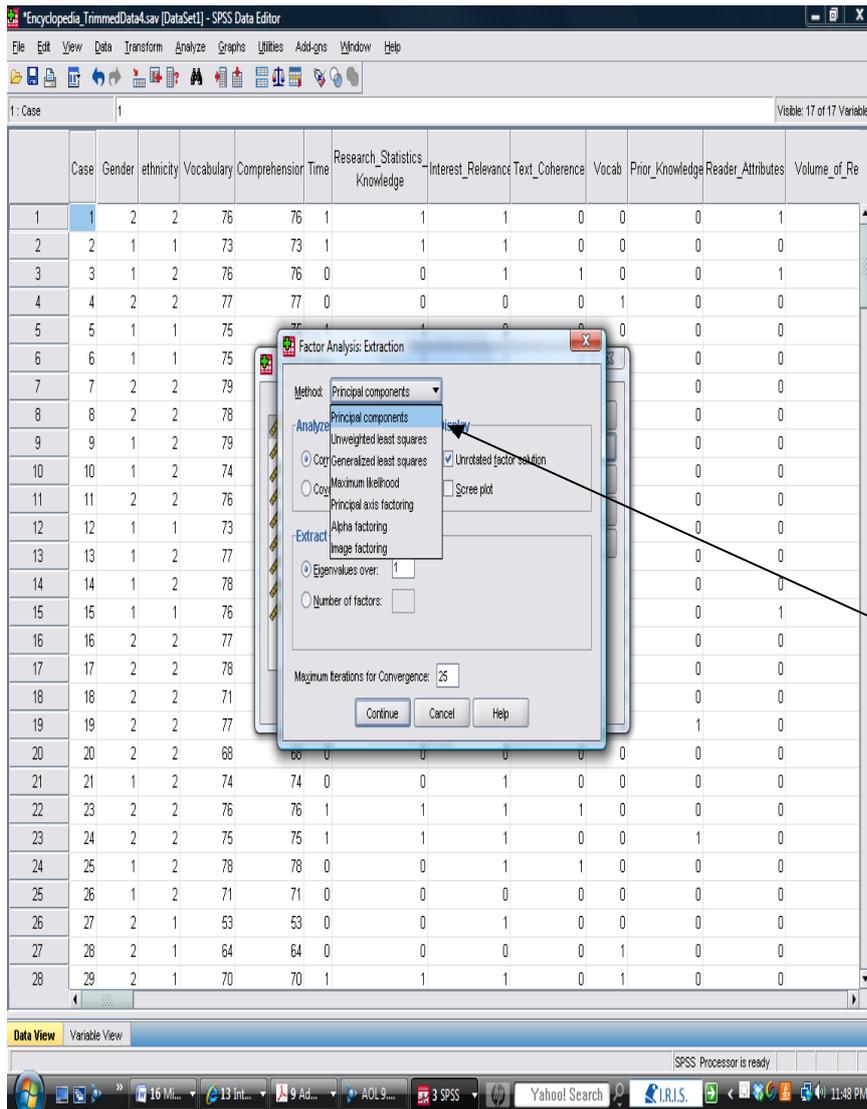


Clicking on the *Extraction* tab provides access for the analyst to select the matrix used for the principal component analysis (i.e., correlation matrix [preferred] vs. covariance matrix), the extraction method, whether the unrotated factor pattern and scree plot are displayed, and the eigenvalue extraction criterion, whether to select the number of factors to be extracted, and the maximum number of iterations for convergence.

Clicking on the *Rotation* tab provides access for the analyst to select whether a varimax rotation or an oblique rotation (e.g., promax) will be used.

Clicking on the *Options* tab provides access for the analyst to select the coefficient display format (e.g., sort by size).

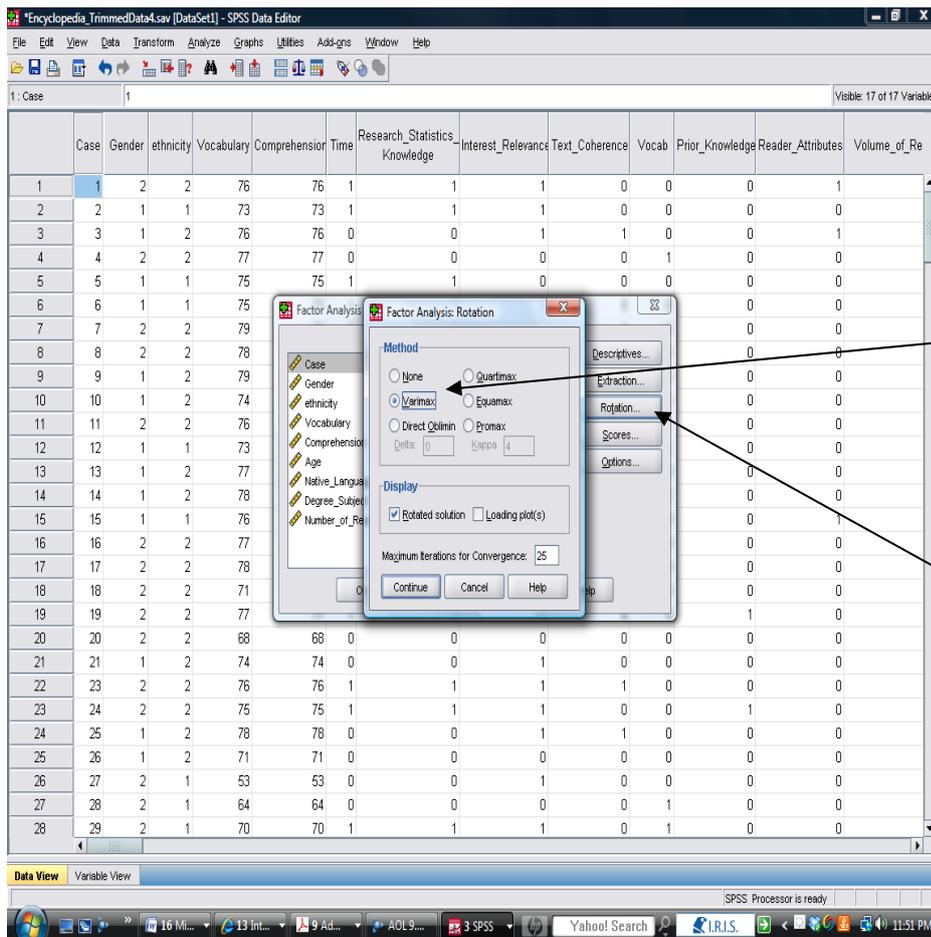
Figure 11. Screenshot showing the eight themes moved over for a principal component analysis



Clicking the *Extraction* tab on the *Factor Analysis* menu produced the *Factor Analysis: Extraction* menu.

SPSS provides seven extraction methods. Benge et al. (2010) used Principal Components analysis.

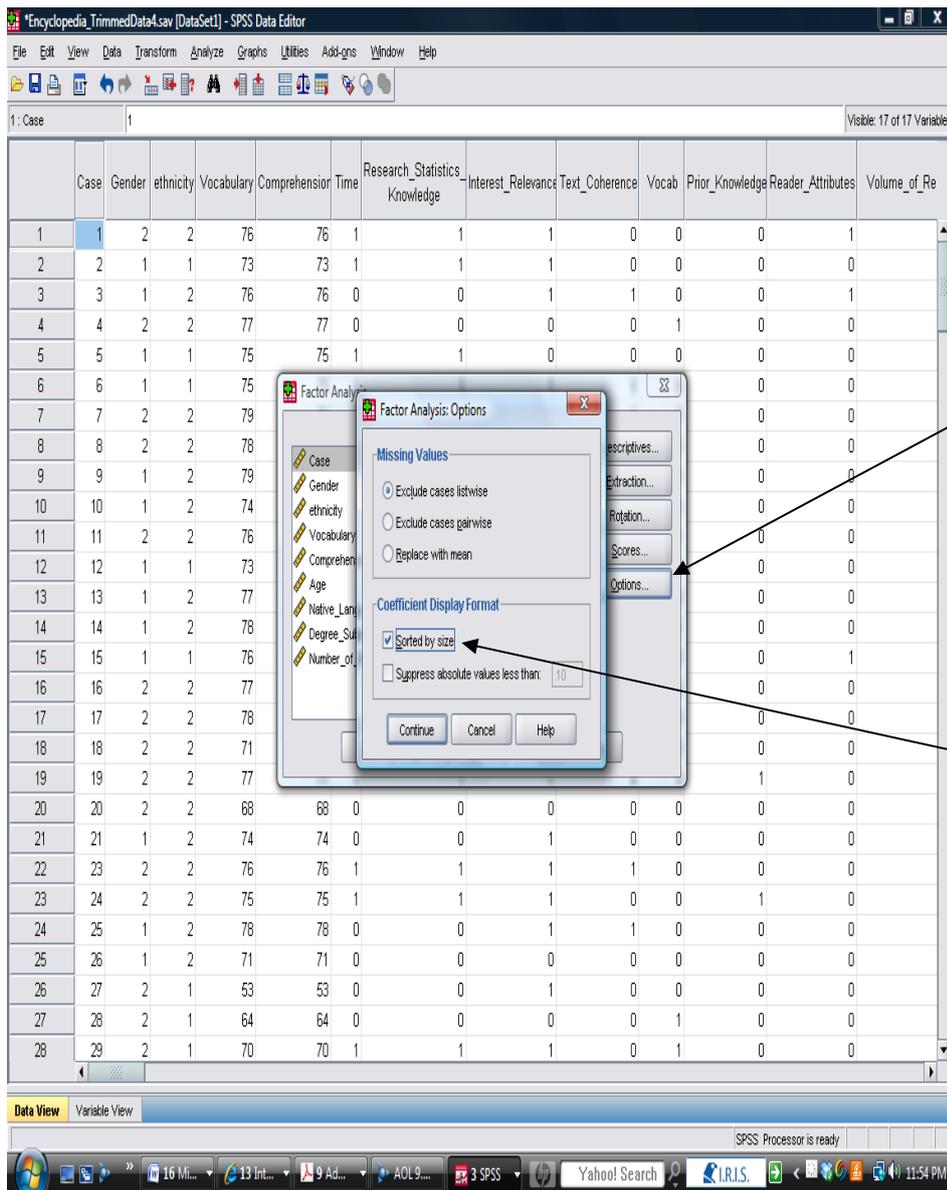
Figure 12. Screenshot showing choice of the Extraction method



Benge et al. (2010) used varimax rotation.

Clicking the *Rotation* tab on the *Factor Analysis* menu produced the *Factor Analysis: Rotation* menu.

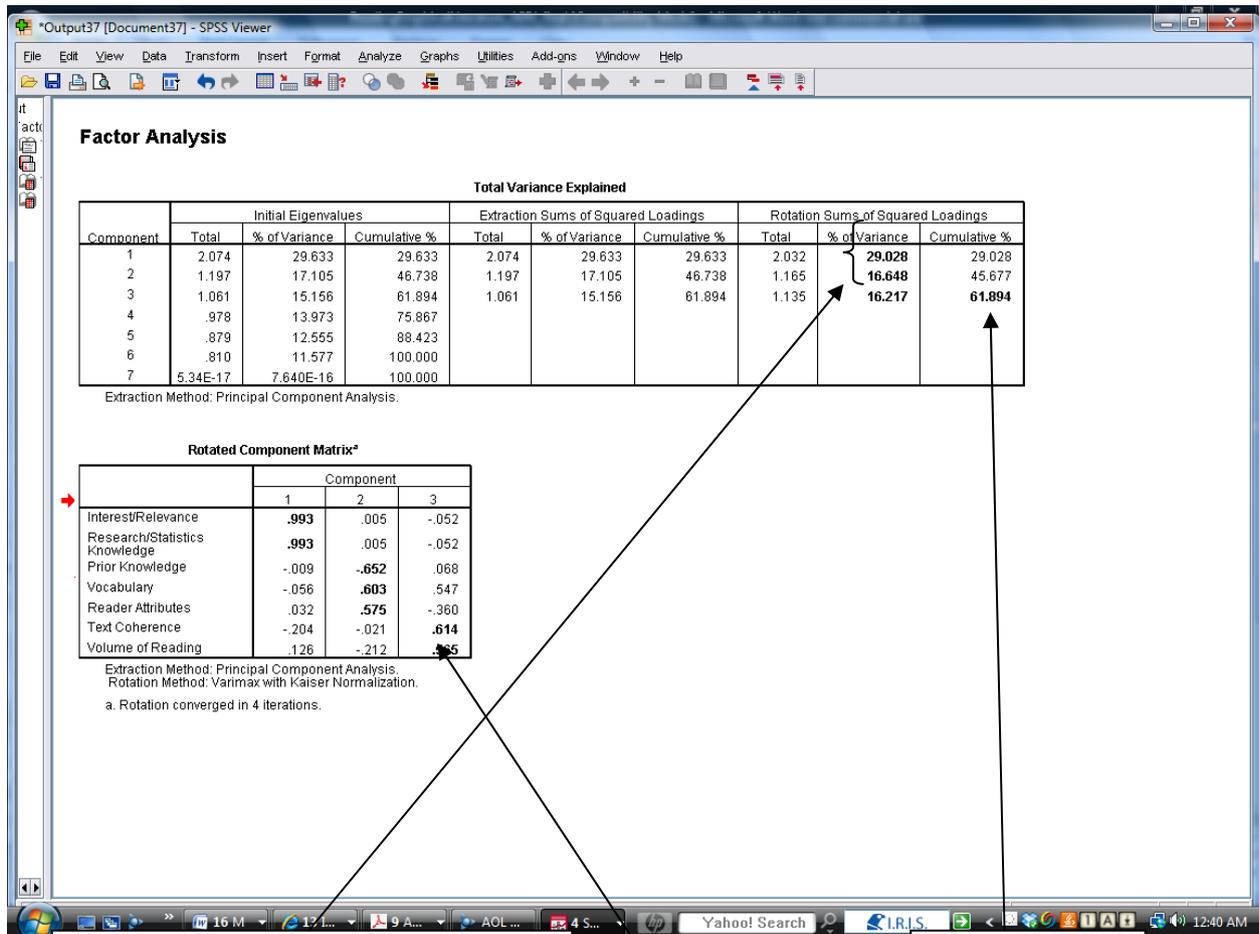
Figure 13. Screenshot showing choice of the Rotation method



Clicking the *Options* tab on the *Factor Analysis* menu produced the *Factor Analysis: Options* menu.

Clicking on the *Sorted by size* tab orders the pattern/structure coefficients by size.

Figure 14. Screenshot showing choice of the Options method



Proportion of variance explained for each of the three extracted factors.

Total proportion of variance explained by the three factors (i.e., 61.894%)

Figures in bold in the *Rotated Component Matrix* denote practically significant pattern/structure coefficients (i.e., > .30) within each of the three extracted and rotated factors.

Figure 15. Screenshot showing output from the principal component analysis

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).