Quality & Quantity (2008) 42:75–87
DOI 10.1007/s11135-006-9037-3

A Framework for Making Quantitative Educational Research Articles More Reader-Friendly for Practitioners

ANTHONY J. ONWUEGBUZIE[1,*], NANCY L. LEECH[2] and JANINE A. WHITCOME[3]
[1]*Department of Educational Measurement and Research, University of South Florida, Tampa, FL, USA;* [2]*University of Colorado at Denver and Health Sciences Center, 1250 14th Str., Denver, CO 80202-1702, USA;* [3]*University of Nevada, 1664 N.Virginia St., Reno, NV 89557-0042, USA*

Abstract. The purpose of the present article is to demonstrate how quantitative research articles can be made much more reader-friendly. In particular, we illustrate how statistical language and research terminology can be simplified in reports. Moreover, using a published article, we demonstrate how quantitative reports could be re-structured to make them more reader-friendly, without sacrificing any important statistical information. We contend that by restructuring these reports, practitioners and stakeholders would be in a much better position to read quantitative research articles, whose findings could then be utilized to improve the quality of education. As such, not only would the divide between researchers and practitioners be reduced, but also educational research studies would have a much bigger impact on schools.

Key words: quantitative research, educational research, quantitative analysis, statistics, quantitative results, reader-friendly articles

1. A Framework for Making Quantitative Educational Research Articles More Reader-Friendly for Practitioners

According to census data, the percentage of elementary and secondary school students retained in grade has risen steadily over the last three decades. A major reason cited for the use of retention is that the targeted student is immature and lagging significantly behind her/his peers academically, emotionally, and/or socially. As surmised by Brooks (2002), a proposed solution is for the child to repeat the same grade level and to be exposed for a second year to the same curriculum, thereby providing the

* Author for correspondence: Anthony J. Onwuegbuzie, Department of Educational Measurement and Research, College of Education, University of South Florida, 4202 East Fowler Avenue, EDU 162, Tampa, FL 33620-7750, USA. E-mail: tonyonwuegbuzie@aol.com

child with an opportunity to mature and to increase their levels of achievement to more grade-appropriate levels (albeit with classmates who are at least one year younger).

Approximately 2.4 million children in the United States are retained each year, costing more than $14 billion dollars and one year of these children's lives (Dawson, 1998; Jimerson, 2001). Unfortunately, a myriad of studies (e.g., Jimerson, 2001) has demonstrated that, for many children, retention represents an ineffective means to improve levels of academic achievement. In particular, in their landmark meta-analytic study, Shepard and Smith (1990) concluded, "Although grade retention is widely practiced, it does not help children to 'catch up.' Retained children do better in the short term, but they are at much greater risk for future failure than their equally achieving, non-retained peers" (p. 84). However, even though evidence against retention was provided as early as in 1990, President Clinton still extolled its virtues by calling for the end of social promotion in his 1998 State of the Union Address (Dawson, 1998). Further, recent educational initiatives aimed at increasing standards and emphasizing accountability, such as *No Child Left Behind*, are likely to lead to increased retention rates (Jimerson, 2001).

For some reason, presidents, politicians, and other policymakers in the United States have not used the findings from these studies on retention. Unfortunately, findings from educational research studies not being used by policymakers is an all too common occurrence. The question to be asked is then "Why are educational research studies largely ignored by policymakers and stakeholders, who are in the best position to use its findings?" We believe that one reason for studies being overlooked stems from the fact that many stakeholders and policymakers find research articles, particularly those representing quantitative investigations, difficult to read. This is ironic, bearing in mind that many applied research articles supposedly are aimed at practitioners. Many quantitative research articles contain statistical jargon that only those who have taken several statistics courses can understand, making them not reader-friendly. In fact, such statistical jargon often induces high levels of anxiety (Onwuegbuzie et al., 1997). Thus, it is likely that stakeholders and policymakers, who may be barely statistically literate, do not read such articles. This, in turn, culminates in many policies being set, such as that relating to school retention, which contradicts the literature base.

Thus, it is clear that measures are needed to make quantitative research articles more reader-friendly. Because the results section tends to be the most difficult part of a research report, we believe that one way of improving the readability of these articles is by reducing the complexity of results section. Below, we use a heuristic example to demonstrate how this might be undertaken.

2. Heuristic Example

In this section, we will demonstrate how to make quantitative research reports more reader-friendly. We will use an article published by Onwuegbuzie et al. (2000) to make this illustration. This article was selected because it utilized a rigorous and systematic multiple regression analysis that included the following 11 components: (a) a check of the analytical assumptions, (b) an explanation of the regression analysis technique used, (c) a description of the effect size indices used, as well as the criteria used for assessing the strength of the relationship, (d) an explanation of the rationale for selecting the final model, (e) a delineation of the statistical and practical significance of the final model selected, (f) a detailed discussion of the checks conducted to assess model adequacy (i.e., analysis of residuals, variance inflation factors, condition numbers), (g) description of the internal replication analysis, (h) delineation of the influence diagnostic analysis, (i) specification of the variables in the final model, and (j) a discussion of the size of the effect pertaining to each independent variable, and (k) a delineation of the structure coefficients. Although all of these 11 components have been deemed as representing good practice by various methodologists (e.g., Onwuegbuzie and Daniel, 2003), readers would not have been exposed to many of these techniques unless they had taken at least three statistics courses. Unfortunately, relatively few practitioners have taken this number of statistics courses. Thus, few consumers of educational research are in a position to understand the results pertaining to all 11 components that were documented in Onwuegbuzie et al.'s (2000) study.

Onwuegbuzie et al. (2000) conducted a study investigating correlates of achievement among students enrolled in several sections of a graduate-level quantitative-based educational research course at a university in the south eastern United States. The theoretical framework for this investigation, though not presented here, can be found by examining the original study. The dependent variable, achievement in the educational research course, was measured using students' course averages. A total of 18 independent variables were examined, comprising cognitive (i.e., academic achievement, study habits, expectation of achievement in educational research course), affective (i.e., research anxiety, composition anxiety, worth of statistics, interpretation anxiety, test and class anxiety, computation self-concept, fear of asking for help, fear of the statistics instructor), and demographic (i.e., age, number of college-level research courses, number of college-level statistics courses, number of college-level mathematics courses, time elapsed since previous college-level math course, course load, students' occupational status) variables. The major analysis undertaken in Onwuegbuzie et al.'s study involved the use of multiple regression. The text excerpt from the results section (excluding the tables and references) is presented in Excerpt 1[a].

Excerpt 1: Results Section from Onwuegbuzie et al. (2000)

Table I presents the correlations between each of the selected independent variables and overall educational research achievement, as well as the means and standard deviations of all variables. Using the Bonferroni adjustment (Maxwell and Delaney, 1990), it can be seen that achievement in educational research correlated negatively with the following variables: worth of statistics, test and class anxiety, computation self-concept, and course load. In addition, achievement correlated positively with study habits, age, and expected overall average for the current educational research course. Table II presents the intercorrelations between all of the predictor variables.

The Shapiro–Wilk test (Shapiro and Wilk, 1965; Shapiro et al., 1968) did not indicate that the distribution of educational research achievement scores was non-normal ($W = 0.97$, $p > 0.05$), thereby justifying the use of multiple regression. In addition, evaluation of assumptions of linearity and homogeneity revealed no threat to multiple regression analysis.

All possible subsets (APS) multiple regression (Thompson, 1995) was used to identify an optimal combination of cognitive, affective, and demographic variables (i.e., independent variables) that predicted achievement in the educational research course. Using this technique, all possible models involving some or all of the independent variables were examined. This method of analysis has been recommended by many statisticians (e.g., Thompson, 1995). Indeed, 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 according to some criterion. For this study, the criterion used was the maximum proportion of variance explained (R^2), which provides an important measure of effect size (Cohen, 1988). An additional index used was Mallow's C_p (Myers, 1986; Sen and Srivastava, 1990).

Squared semi-partial correlation coefficients, also known as part correlations, represent the amount by which R^2 is reduced if a particular independent variable is removed from the regression equation. That is, squared semi-partial correlation coefficients express the unique contribution of the independent variable as a proportion of the total variance of the dependent variable (Cohen, 1988). Similarly, squared partial correlation coefficients represent the unique contribution of the independent variable as a proportion of R^2. In this study, squared partial correlation coefficients, like R^2, were used directly as effect size estimates, as recommended by Cohen (1988). According to Cohen (1988), for multiple regression models in the behavioral sciences, squared partial correlation values between 2 and 12.99% suggest small effect sizes, values between 13 and 25.99% indicate medium effect sizes, and values of 26% and greater suggest large effect sizes. These same criteria were used to assess whether the proportion of variance explained by the independent variables, R^2, was suggestive of a small, medium, or large effect.

Table III presents the unstandardized regression coefficients and intercept, the standard error of the unstandardized coefficients, the standardized regression coefficients, the structure coefficients, the squared semi-partial correlations, the squared partial correlation coefficients, and the squared multiple correlation coefficient (R^2) of the chosen model, as well as tolerance statistics, variance inflation factors, and condition numbers. The APS multiple regression analysis revealed that a model containing eight variables provided the best fit to these data. In fact, the *best* nine-variable model, in which the number of statistics courses taken was added to the model, only increased the proportion of variance explained by 1.6%. In addition, Mallow's C_p was closer in value to the number of regressor variables (Myers, 1986; Sen and Srivastava, 1990) with the eight-variable solution than with any nine-variable solution.

The selected model indicated that the following eight variables contributed significantly ($F[8, 112] = 14.26$, $p < 0.0001$) to the prediction of educational research achievement: research anxiety, computation self-concept, study habits, age, course load, number of college-level research courses taken, expectation of educational research achievement, and grade point average (Table III). These eight variables combined to explain 50.5% of the variation in educational research achievement. Using Cohen's (1988) criteria for assessing the predictive power of a set of independent variables in a multiple regression model, the proportion of variance explained indicates a large effect size, because it well exceeds 26%.

An inspection of the studentized residuals generated from the model (Myers, 1986) suggested that the assumptions of normality, linearity, and homoscedasticity were met. Using the Bonferroni adjustment, none of the studentized residuals suggested that outliers were present. Additionally, an examination of the tolerance statistics, the variance inflation factors, and the condition numbers of the selected regression model indicated strongly that no multicollinearity was present. Specifically, all variance inflation factors (Table III), which indicate the extent to which the variance of an individual regression coefficient has been inflated by the presence of collinearity (Sen and Srivastava, 1990), are much less than 10, which is Myer's (1986) criteria for suspecting the presence of multicollinearity. Indeed, all the variance inflation factors were relatively close to unity, which indicates no relationship among the regressor variables. Condition numbers represent the ratio of the largest to the smallest eigenvalues, which, in turn, are measures of the strength of linear dependency among regressor variables. From Table III, it can be seen that all condition numbers are much less than Myer's (1986) cut-off score of 1000, again suggesting that multicollinearity is not a feature of these data.

As recommended by Thompson (1994), several empirical internal replicability analyses were conducted to evaluate further the adequacy of the selected regression model. Specifically, a jackknife method was used (Crask and Perreault, 1977). This involved conducting 121 separate regression analyses (each fitting the eight-variable model), whereby each analysis involved dropping the ith participant until every subject had been eliminated exactly once. That is, each of the resultant 121 regression models utilized 120 subjects (i.e., $n-1$ subjects, where $n=$ the total sample size). The 121 adjusted and unadjusted R^2 values which were generated from these models were examined for stability. The summary statistics pertaining to this analysis are presented in Table IV. Assuming that the sample estimates of the multiple correlation coefficients are normally distributed (as suggested by the closeness of the mean and median values for both adjusted and unadjusted estimates), it can be seen that the 95% confidence interval about the parameter estimate lies between 50.4 and 50.6% for the unadjusted R^2 and between 46.8 and 47.0% for the adjusted R^2. Encouragingly, these intervals are not only very narrow, but they contain the estimates calculated using the complete data (i.e., $R^2=50.5\%$, adjusted $R^2=46.9\%$; c.f., Table III) – suggesting that neither the adjusted R^2 nor the unadjusted R^2 are impacted by variations in the sample.

Finally, the following additional influence diagnostics were examined: (1) the number of estimated standard errors (for each regression coefficient) that the coefficient changes if the ith observation were set aside (i.e., DFBETAS); (2) the number of estimated standard errors that the predicted value changes if the ith point is removed from the data set (i.e., DFFITS); and (3) the reduction in the estimated generalized variance of the coefficient over what would have been produced without the ith data point (i.e., COVRATIO). Using criteria recommended in the literature (e.g., Myers, 1986; Sen and Srivastava, 1990), no subject generated DFBETAS, DFFITS, or COVRATIO values that were large enough to indicate that (s)he represented an outlying observation – again suggesting sample invariance.

> The regression model suggests that students with the lowest levels of performance in educational research courses tended to have at least one of the following eight characteristics: younger, lower overall academic achievers, higher levels of research anxiety, higher levels of statistics anxiety associated with computation self-concept, poorer study habits, lower expectations for their overall achievement in the educational research course, more previous research methodology courses, and heaviest course loads.
>
> From the squared semi-partial coefficients (Table III), it can be seen that computation self-concept and students' expectations of their achievement were the best predictors of overall achievement, each explaining 12.2% of the variance. The squared partial coefficients for these variables (i.e., 19.7 and 19.8%, respectively) suggest a moderate effect size. These variables were followed, respectively, by age, study habits, number of college-level research courses, research anxiety, course load, and grade point average. The predictive power of these latter variables represented small effect sizes. An examination of the structure coefficients (Table III), using a cutoff correlation of 0.3 recommended by Lambert and Durand (1975) as an acceptable minimum loading value, suggests that all eight variables made important contributions to the model (even grade point average, which explained the smallest proportion of variance). The fact that both the standardized and structure coefficients pertaining to all variables were noteworthy indicates that none of these constructs acted as suppressor variables (Thompson, 1998; Thompson and Borello, 1985, pp. 57–61).

Critique of Excerpt 1

As can be seen from Excerpt 1, the results section of Onwuegbuzie et al.'s investigation is very technical. Consequently, it is unlikely that practitioners and others students who do not specialize in statistics would be able to understand much of this section. This lack of understanding likely would demotivate them from reading the results section, even though this section contains some direct information about the findings. Even more disturbingly, the complexity of the results section might even lead to them not reading any part of the article at all, because they might assume that the whole study is similarly too complex for them to read. Regardless of whether the remainder of the article is read, not reading the results section could be problematic because the reader is forced to accept the interpretations of the researcher that follow in the discussion section.

We contend that the vast majority of the technical components contained in results section of quantitative studies does not need to be presented there. We recognize that this information is important, and its removal would make it difficult for journal reviewers to critique the analytical methodology used, as well as to assess the consistency between analysis and results. Notwithstanding, we believe that immersing or interspersing the technical details with the results that directly address the research questions and/or test the study hypotheses is likely to lead to avoidance behaviors on the part of the untrained reader. Moreover, we believe that most of the technical information should be moved to the appendix section of the article. With this in mind, the results section of Onwuegbuzie et al.'s study could be drastically reduced, as illustrated in Excerpt 2.

Excerpt 2: Abridged suggested change to the Results Section of Onwuegbuzie et al. (2000)

> Table I presents the correlations between each of the selected independent variables and overall educational research achievement, as well as the means and standard deviations of all variables. The highlighted correlations in this table are statistically significant. Examination of these highlighted coefficients revealed that achievement in educational research correlated negatively with the following variables: worth of statistics, test and class anxiety, computation self-concept, and course load. In addition, achievement correlated positively with study habits, age, and expected overall average for the current educational research course. Table II presents the intercorrelations between all of the predictor variables.
>
> A multiple regression analysis was used to determine which of the 11 selected predictor cognitive, affective, and demographic variables (i.e., independent variables) predicted achievement in the educational research course.[1,2,3] Table III presents the eight variables that significantly predicted educational research achievement.[4,5,6] These variables were: research anxiety, computation self-concept, study habits, age, course load, number of college-level research courses taken, expectation of educational research achievement, and grade point average.[7,8,9,10]
>
> The regression model suggests that students with the lowest levels of performance in educational research courses tended to have at least one of the following eight characteristics: younger, lower overall academic achievers, higher levels of research anxiety, higher levels of statistics anxiety associated with computation self-concept, poorer study habits, lower expectations for their overall achievement in the educational research course, more previous research methodology courses, and heaviest course loads. Table III also reveals that computation self-concept and students' expectations of their achievement were the best

predictors of overall achievement because they each explained 12.2% of the variance, which is larger than any other varable.[11,12]

All the technical information removed from the results section could then be moved to the Appendix section. as demonstrated in Excerpt 3. This excerpt contains 12 footnotes. Any of these footnotes could be referred to again in the discussion section.

Excerpt 3: Suggested Appendix for the Onwuegbuzie et al. (2000) study

---

1. The Shapiro–Wilk test (Shapiro and Wilk, 1965; Shapiro et al., 1968) did not indicate that the distribution of educational research achievement scores was non-normal ($W = 0.97$, $p > 0.05$), thereby justifying the use of multiple regression. In addition, evaluation of assumptions of linearity and homogeneity revealed no threat to multiple regression analysis.
2. Specifically, an all possible subsets (APS) multiple regression (Thompson, 1995) was used to identify an optimal combination of cognitive, affective, and demographic variables (i.e., independent variables) that predicted achievement in the educational research course.[1] Using this technique, all possible models involving some or all of the independent variables were examined. This method of analysis has been recommended by many statisticians (e.g., Thompson, 1995). Indeed, 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 according to some criterion. For this study, the criterion used was the maximum proportion of variance explained (R^2), which provides an important measure of effect size (Cohen, 1988). An additional index used was Mallow's C_p (Myers, 1986; Sen and Srivastava, 1990).
3. Squared semi-partial correlation coefficients, also known as part correlations, represent the amount by which R^2 is reduced if a particular independent variable is removed from the regression equation. That is, squared semi-partial correlation coefficients express the unique contribution of the independent variable as a proportion of the total variance of the dependent variable (Cohen, 1988). Similarly, squared partial correlation coefficients represent the unique contribution of the independent variable as a proportion of R^2. In this study, squared partial correlation coefficients, like R^2, were used directly as effect size estimates, as recommended by Cohen (1988). According to Cohen (1988), for multiple regression models in the behavioral sciences,

squared partial correlation values between 2 and 12.99% suggest small effect sizes, values between 13 and 25.99% indicate medium effect sizes, and values of 26% and greater suggest large effect sizes. These same criteria were used to assess whether the proportion of variance explained by the independent variables, R^2, was suggestive of a small, medium, or large effect.

4. Table III presents the unstandardized regression coefficients and intercept, the standard error of the unstandardized coefficients, the standardized regression coefficients, the structure coefficients, the squared semi-partial correlations, the squared partial correlation coefficients, and the squared multiple correlation coefficient (R^2) of the chosen model, as well as tolerance statistics, variance inflation factors, and condition numbers.

5. The selected eight-variable model was statistically significant ($F[8, 112] = 14.26$, $p < 0.0001$).

6. In fact, the best nine-variable model, in which the number of statistics courses taken was added to the model, only increased the proportion of variance explained by 1.6%. In addition, Mallow's C_p was closer in value to the number of regressor variables (Myers, 1986; Sen & Srivastava, 1990) with the eight-variable solution than with any nine-variable solution.

7. These eight variables combined to explain 50.5% of the variation in educational research achievement. Using Cohen's (1988) criteria for assessing the predictive power of a set of independent variables in a multiple regression model, the proportion of variance explained indicates a large effect size, because it well exceeds 26%.

8. An inspection of the studentized residuals generated from the model (Myers, 1986) suggested that the assumptions of normality, linearity, and homoscedasticity were met. Using the Bonferroni adjustment, none of the studentized residuals suggested that outliers were present. Additionally, an examination of the tolerance statistics, the variance inflation factors, and the condition numbers of the selected regression model indicated strongly that no multicollinearity was present. Specifically, all variance inflation factors (Table III), which indicate the extent to which the variance of an individual regression coefficient has been inflated by the presence of collinearity (Sen and Srivastava, 1990), are much less than 10, which is Myer's (1986) criteria for suspecting the presence of multicollinearity. Indeed, all the variance inflation factors were relatively close to unity, which indicates no relationship among the regressor variables. Condition numbers represent the ratio of the largest to the smallest eigenvalues, which, in turn, are measures of the strength of linear dependency among regressor variables. From Table III, it can be seen that all condition numbers

are much less than Myer's (1986) cut-off score of 1000, again suggesting that multicollinearity is not a feature of these data.
9. As recommended by Thompson (1994), several empirical internal replicability analyses were conducted to evaluate further the adequacy of the selected regression model. Specifically, a jackknife method was used (Crask and Perreault, 1977). This involved conducting 121 separate regression analyses (each fitting the eight-variable model), whereby each analysis involved dropping the ith participant until every subject had been eliminated exactly once. That is, each of the resultant 121 regression models utilized 120 subjects (i.e., n-1 subjects, where n = the total sample size). The 121 adjusted and unadjusted R^2 values which were generated from these models were examined for stability. The summary statistics pertaining to this analysis are presented in Table IV. Assuming that the sample estimates of the multiple correlation coefficients are normally distributed (as suggested by the closeness of the mean and median values for both adjusted and unadjusted estimates), it can be seen that the 95% confidence interval about the parameter estimate lies between 50.4 and 50.6% for the unadjusted R^2 and between 46.8 and 47.0% for the adjusted R^2. Encouragingly, these intervals are not only very narrow, but they contain the estimates calculated using the complete data (i.e., $R^2 =$ 50.5%, adjusted $R^2 =$46.9%; c.f. Table III) – suggesting that neither the adjusted R^2 nor the unadjusted R^2 are impacted by variations in the sample.
10. The following additional influence diagnostics were examined: (1) the number of estimated standard errors (for each regression coefficient) that the coefficient changes if the ith observation were set aside (i.e., DFBETAS); (2) the number of estimated standard errors that the predicted value changes if the ith point is removed from the data set (i.e., DFFITS); and (3) the reduction in the estimated generalized variance of the coefficient over what would have been produced without the ith data point (i.e., COVRATIO). Using criteria recommended in the literature (e.g., Myers, 1986; Sen and Srivastava, 1990), no subject generated DFBETAS, DFFITS, or COVRATIO values that were large enough to indicate that (s)he represented an outlying observation–again suggesting sample invariance.
11. From the squared semi-partial coefficients (Table III), it can be seen that computation self-concept and students' expectations of their achievement were the best predictors of overall achievement, each explaining 12.2% of the variance. The squared partial coefficients for these variables (i.e., 19.7 and 19.8%, respectively) suggest a moderate effect size. These variables were followed, respectively, by age, study habits, number of college-level research courses, research anxiety,

course load, and grade point average. The predictive power of these latter variables represented small effect sizes.

12. An examination of the structure coefficients (Table III), using a cutoff correlation of 0.3 recommended by Lambert and Durand (1975) as an acceptable minimum loading value, suggests that all eight variables made important contributions to the model (even grade point average, which explained the smallest proportion of variance). The fact that both the standardized and structure coefficients pertaining to all variables were noteworthy indicates that none of these constructs acted as suppressor variables (Thompson and Borello, 1985; Thompson, 1998).

3. Summary and Conclusions

Quantitative researchers heavily rely on numbers to convince readers to accept their findings as scientifically valid (Sandelowski, 2003). Many of these researchers are torn between writing a report that reaches a wide audience, particularly stakeholders and policymakers, and writing an article that enhances their reputations as scholars. Unfortunately, in attempting to fulfill the latter goal, the former goal appears to have suffered, and the field of education is permeated by many articles that can only be read by an elite few. Thus, it should not be surprising that research articles are often ignored by stakeholders and policymakers when making educational decisions. Thus, the present article set out to demonstrate how quantitative research articles can be made much more reader-friendly. We believe that making articles more reader-friendly would increase the impact that quantitative studies can have on educational policy by improving the chances that they would be read by "those who count."

Note

[a]Reprinted with kind permission of the Mid-South Educational Research Association and the Editors of Research in the Schools.

References

Brooks, R. (2002). School retention: A common practice but is it effective? Retrieved May 15, 2006, from http://www.cdl.org/resources/reading_room/print/retention.html

Dawson, P. (1998). A primer on student grade retention: What the research says. *Communiqué* 26: 28–30.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review* 30: 420–437.

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). Available: http://cie.ed.asu.edu/volume6/number2/

Onwuegbuzie, A. J., DaRos, D. & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics* 19: 11–35.

Onwuegbuzie, A. J., Slate, J., Paterson, F., Watson, M. & Schwartz, R. (2000). Factors associated with underachievement in educational research courses. *Research in the Schools* 7(1): 53–65.

Sandelowski, M. (2003). Tables or tableaux? The challenges of writing and reading mixed methods studies. In: A. Tashakkori & C. Teddlie (eds.), *Handbook of Mixed Methods in Social and Behavioral Research* . Thousand Oaks, CA: Sage, pp. 321–350

Shepard, L. A. & Smith, M. L. (1990). Synthesis of research on grade retention. *Educational Leadership* 47(8): 84–88.