Multiple Linear Regression: A Return to Basics in Educational Research

Winona Burt Vesey
University of Houston, Clear Lake

Jermaine T. Vesey
University of Texas, San Antonio

Antionette D. Stroter
Liberty University

Kyndra V. Middleton
Howard University

A content analysis of the *American Educational Research Journal* and the *Educational Evaluation and Policy Analysis* journal for the use of multiple linear regression (MLR) was conducted. Two hundred articles were randomly sampled and coded to determine if basic reporting guidelines were followed. Results showed that standard reporting methods for MLR were not followed and the use of stepwise MLR was on the rise. Manuscripts using MLR did not apply necessary corrections for inflated Type I error rates. The majority of the sampled articles did not include key summary statistics, which violated the American Educational Research Association's principle of transparency. The lack of consistency in reporting hindered critique of work, meta-analysis, and theory development.

Multiple linear regression (MLR) is a common statistical technique used in educational research (Elmore & Woehlke, 1996). Used in experimental and non-experimental research designs alike, MLR involves the use of one or more predictor (i.e., independent) variables predicting some criterion (i.e., dependent) variable. Research conclusions based on MLR are used to influence education policy decisions (Clements & Sarama, 2008; Ingersoll, 2001; Stern, Dayton, Paik, & Weisberg, 1989), inform school reform efforts (Desimone, Smith, Baker, & Uano, 2005), determine variables considered in college admission (e.g., Zwick & Sklar, 2005), and identify relationships between school climate and student achievement (e.g., McInerney, Roche, McInerney, & Marsh, 1997; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). Given the pervasive use of this technique in educational research, the improper interpretation of MLR results are far reaching. Hierarchical MLR (HMLR) and stepwise MLR (SMLR) are particularly susceptible to misreporting. Regrettably, as is detailed in this study, many educational researchers fail to avoid common misinterpretations of HMLR and SMLR. Research and statistical analysis classes typically focus on the scientific method and its fundamental guidance to conducting good research. Specifically, theory is developed and tested to determine if it can be supported. When theories endure multiple tests they are regarded as robust and acceptable – statistical analysis is a critical component in this process. In this respect, replication and statistical validity are critical to the development of sound theory. If researchers are unable to replicate research due to the lack of standardized reporting, or if researchers are concluding that hypotheses are supported when in fact they are not; proper development and acceptance of theory is jeopardized.

To facilitate replication of studies, transparency in reporting is essential. Transparency in research findings refers to the practice of revealing the key methodological components necessary for scrutiny and replication of research findings (American Educational Research Association (AERA), 2006). Transparency allows researchers to critique the statistical validity of a study, conduct replication studies, and ultimately ensure the accuracy of research claims. Despite the known importance of standard reporting practices and statistical validity, many researchers have failed to provide adequate transparency in their analysis. The purpose of this paper is to call attention to the frequency of inaccurate statistical claims (i.e., specifically as related to regression), call for statistical reporting standards, and provide recommendations to aid in the development of said standards.

.

Importance of Study

The use of MLR is common among education researchers (Elmore & Woehlke, 1996). In Elmore and Woehlke's review of articles published in the *American Educational Research Journal* (AERJ), *Educational Researcher* (ER), and the *Review of Educational Research* (RER) from 1988 to 1995, MLR/correlation emerged as the third most used statistical method following analysis of variance/analysis of covariance and descriptive methods (1996). MLR continues to be utilized in contemporary research. In our random sample of 200 articles, 35% of our articles were published from 1968 to 1995. The remaining 65% of the articles were published between 1996 and 2008, supporting the pervasiveness of the method in recent years. Given the popularity of MLR (Elmore & Woehlke), it is important to establish standardized reporting conventions and ensure that MLR results are reported accurately. In a recent article, Zientek and Thompson (2009) called for at a minimum, the reporting of matrix summaries when using continuous data (e.g., correlation matrix and standard deviations, or the variance-covariance matrices) as matrices support and encourage meta-analytic thinking. In an earlier paper, Thompson (2007) also called for basic reporting and research standards. The following quote from Thompson reflects his belief in the value of creating statistical reporting standards that reflect transparency and enable replication: "Vital aspects of scholarship include exposing one's conclusions and their warrants to public scrutiny and disseminating one's findings" (Thompson, 2007, p. 18).

Standards for Reporting Empirical Research

Having standards in reporting statistical findings is important for a variety of reasons, namely it makes it easier to generalize findings across fields, ensures accuracy, enables understanding, provides information necessary for replication, and allows researchers to conduct meta-analysis. Our call for reporting standards is shared by other researchers and professional organizations. For example, both AERA and the American Psychological Association (APA) provide guidelines that encourage reporting standards and transparency. Specifically, AERA (2006) outlines two overarching principles for reporting empirical research: sufficiency and transparency. In short, adequate evidence should be provided to support results and conclusions and reports should be transparent. "Reporting that takes these principles into account permits scholars to understand one another's work, prepares that work for public scrutiny, and enables others to use that work" (AERA, 2006, p. 33). AERA provides further detail for the area of analysis and interpretation reporting. Specifically, for quantitative methods, AERA calls for a statement of statistical analyses and why they were appropriate; descriptive and inferential statistics; discussion of considerations that arose during data collection and processing such as, missing data or attrition; considerations identified as a result of data analysis (e.g., violations of assumptions); and inclusion of a measure of effect size for each statistical result; standard error or confidence interval; test statistics and its significance level for hypothesis testing; and a qualitative description of the index of the effect.

The APA (2010) publication manual also provides guidelines for standard reporting practices. Specifically, they call for summary descriptive data, variance-covariance or correlation matrices, and results of inferential statistics (e.g., observed values, degrees of freedom, p values, standard errors, and effect sizes).

Conventions for reporting analysis of variance, t-tests, and correlation results have been well established for many years (Daniel, 2001; Schafer, 1991). Nonetheless, similar standards for reporting MLR results have not been established (Courville & Thompson, 2001; Schafer). In his 1991 editorial, Schafer offers recommendations for reporting hierarchical regression results. Schafer proposes that authors report both descriptive and inferential statistics (e.g., correlation matrix for the predictors, df column, R^2 change column, and p values for each F ratio), and predictors listed in the order of their inclusion in the analysis. Schafer emphasizes reporting results with sufficient information in an interpretable way without losing the ability for replication, thereby highlighting the importance of transparency in research. Schafer further notes that if exact p values are reported, then it is not as important to indicate which are below the alpha level set by the researcher, but simply indicate the *a priori* alpha level in the text. Near the end of the article, Schafer states "Whether this or some other format becomes popular remains to be seen, but it seems clear that some conventional way to report multiple regression outcomes is needed" (Schafer, p. 3). As evidenced in the content review below, conventional reporting for MLR has yet to take hold. Moreover, a standard for identifying the type of MLR and proper adjustments for experimentwise error rates were absent in most cases, resulting in inflated Type I error rates and inaccurate statistical claims.

Multiple Regression Typology

Multiple regression is the process of predicting a dependent, outcome variable from a set of independent, predictor variables. The dependent and independent variables can take several forms: continuous, dichotomous, or polytomous, to name a few. The level of measurement of the dependent variable (DV) is typically used to define the multiple regression technique. For example, in logistic regression the dependent variable is dichotomous; whereas in MLR the dependent variable is continuous.

Regression techniques are further classified by the procedures used to enter the independent variables (IV) to obtain the final equation. Simultaneous MLR is when all independent variables are entered into the regression equation at one time, which results in one hypothesis being tested. Hierarchical multiple linear regression (HMLR) is when the entry order of the independent variables is predetermined by the researcher and there are two or more stages of variable entry into the regression equation. For example, a researcher might first enter demographic variables as control variables in stage one and then enter a second set of variables in stage two, and focus on a change in R^2 (i.e., the amount of variance in the DVs accounted for by the IVs) at each stage. As a second example, the researcher might enter a single IV, such as self-efficacy, followed by another single variable, such as school climate, at stage two. This process would continue for all variables the researcher chooses to enter. Lastly, stepwise regression (also known as empirical multiple regression) is when statistical software determines the entry order of IVs based on which variables contribute most to prediction at a given step in the regression equation (Hoyt, Leierer, & Millington, 2006).

For the purposes of this study, we have categorized multiple regression techniques using the following terms: simultaneous multiple regression, hierarchical multiple regression, and stepwise multiple regression (Hoyt et al., 2006). Methodologists in the psychological community have recommended that SMLR be used rarely, or not at all, in academic research (Cohen, Cohen, West, & Aiken, 2003; Thompson, 1995). The primary reason was that stepwise procedures yield data-dependent results that are unlikely to generalize to future samples (Hoyt et al.). These authors take a similar stance and recommend that researchers never use the stepwise method in education or any other discipline. A computer program is not sufficient to determine the importance of a variable; instead the literature and theory should guide decisions. Additionally, statistical conclusion validity, as discussed in the next section, is often violated when conducting stepwise regression. Furthermore, as demonstrated in our content analysis, neither hierarchical nor stepwise regressions are properly identified in education research and results are often misreported.

Compromising Statistical Conclusion Validity

Statistical conclusion validity refers to the accuracy of a conclusion regarding the relationship between variables (Shaddish, Cook, & Campbell, 2002). In MLR, statistical conclusion validity is often violated when researchers fail to properly adjust alpha levels to compensate for multiple hypothesis testing. When researchers use HMLR or SMLR to test multiple combinations of variables, they are in fact testing multiple hypotheses. Each variable or set of variables entered into and removed from the regression equation represents a separate hypothesis test. When using stepwise regression procedures (e.g., forward and backward regression), the software executes the adding and removing of the variables and provides the model with the best R^2 to the researcher. This is particularly dangerous if the researcher is unaware of the exact number and order of steps used by the software to derive the final model and further supports the authors' position to never use SMLR.

Using hierarchical and stepwise regression and not adjusting the alpha level is the same as testing multiple hypotheses while holding the alpha level constant. The reason for the adjustment of the alpha level is the same reason that researchers conduct an ANOVA when there are three or more groups being compared. If three separate *t* tests are conducted, the result of this practice is that the researcher is testing the hypothesis at an inflated Type I error rate, which could result in variables being identified as "significant" predictors when they are not. Additionally, running simultaneous regression and not adjusting the alpha level produces a similar problem as hierarchical and stepwise regression in that multiple tests are performed on the same sample data in an attempt to find the best predictors. Not adjusting for the multiple tests can inflate Type I error if several different predictors are tested. In the following section, we provide a more thorough explanation of the experimentwise error rate problem.

Error Rates

To begin, we offer a brief reminder of the difference between alpha levels, *p* values, and error rates. The alpha level is the standard set by the researcher before statistical tests are conducted. Alpha levels are commonly set at .05, .01, and .001 in education research and determine the probability of obtaining a sample mean in the critical region when the null hypothesis is true. In other words, the alpha level controls the risk of making a mistake or a Type I error (Gravetter & Wallnau, 2007). Ultimately then, the risk of a Type I error is in the control of the researcher. The *p* value, or probability value, is related to the test statistic and defines the probability of observing the sample results actually obtained, given that the null hypothesis is true. When the *p* value is equal to or less than the alpha level, the null hypothesis is rejected. Lastly, the error rate, or Type I error (synonymous with the alpha level), is also defined as the probability of rejecting a null hypothesis when in fact it is true (Salkind, 2007).

When conducting multiple hypothesis tests in an experiment, you have both a testwise error rate and an experimentwise error rate. Testwise error rate is the probability of making a Type I error in a single hypothesis test and should be set by researchers *a priori*. Experimentwise error rate, also known as familywise error rate, is the probability of having made a Type I error within a set of hypothesis tests (Thompson, 1995). The experimentwise error rate is inflated for every hypothesis tested on a single set of data in a given experiment (Altman, 2000). Each hypothesis test conducted can be considered as a separate experiment.

Experimentwise Type I error rate is affected by the number of tests (hypotheses) ran using a single sample (Thompson, 1995). When conducting multiple hypothesis tests, the inflated experimentwise error rate (α_{ew}) can be calculated using the Bonferroni inequality (Love, 1988):

$$\alpha_{ew} \leq 1 - (1 - \alpha_{Tw})^k \qquad (1)$$

where k is the number of perfectly uncorrelated hypotheses being tested and α_{Tw} is the testwise alpha level (Altman, 2000). As an example, if you have three different models with variables being entered separately, an alpha initially set at .05 becomes an alpha of .14 using the Bonferroni inequality.

Mundfrom, Perrett, Schaffer, Piccone, and Roozeboom (2006) further propose that when unadjusted *t* tests are used for individual variable selection in simultaneous linear regression, Type I error is even further affected. Researchers using unadjusted alpha levels exponentially inflate the Type I error rate depending on the number of independent variables in the model and the number of independent variables that are correlated with the dependent variable (Mundfrom et al.). As a result, Type I errors are committed, which means variables are identified as “significant” predictors when in fact they may not be. Mundfrom et al. suggest that when conducting multiple hypothesis tests, researchers should control the testwise error rate by using the Bonferroni correction (Altman, 2000):

$$\alpha_{Tw}^{*} = \alpha_{Tw/k} \qquad (2)$$

where k is the number of hypothesis tests being conducted and α_{Tw} is the testwise error rate. Roozeboom, Mundfrom, and Perrett (2008) later developed a modified Bonferroni correction in an effort to maintain greater statistical power

$$\alpha_{Tw}^{*} = \alpha_{Tw/k(1-q)} \qquad (3)$$

where the numerator remains the same nominal alpha value as in equation 2, but the denominator becomes the number of tests performed (k), multiplied by one minus the proportion of nonzero relationships between the dependent and independent variables.

In general, the Bonferroni correction (also known as the Dunn test) adjusts the inflated experiment-wise alpha level by dividing the original testwise error rate (α_{Tw}) by the number of hypotheses being tested (k) yielding a new testwise error rate (α_{Tw}^{*}). Consequently each hypothesis (or *post hoc*) test uses the new testwise error rate to keep the experimentwise error rate at the appropriate level. For example, if there are three comparisons made with an overall alpha level of .05, each comparison would be held to an alpha level of .02 (i.e., .05/3 = .02), thereby maintaining the experiment wise error rate of .05 (Gravetter & Wallnau, 2007). Below we mention additional alternatives to the traditional Bonferroni correction that purport to have greater power than Bonferroni’s correction yet maintain its flexibility for use with tests such as MLR and correlations.

Sidak-Bonferroni

Sidak (1967) suggested a modification of the Bonferroni formula that would have less impact on statistical power than the Bonferroni method and retain much of its flexibility (Keppel & Wickens, 2004). Instead of dividing by the number of comparisons, there is a slightly more complicated formula:

$$\alpha_{S\text{-}B} = 1 - (1 - \alpha_{FWE})^{1/c} \qquad (4)$$

where $\alpha_{S\text{-}B}$ is the Sidak-Bonferroni alpha level used to determine statistical significance (a value less than .05), α_{FWE} is the computed testwise error according to Formula 1, and *c* is the number of comparisons or statistical tests conducted in the study. The *p* values obtained from the results of the analysis must be smaller than $\alpha_{S\text{-}B}$ to be considered significant (Olejnik, Li, Supattathum, & Huberty, 1997).

Methodology

Data Sources and Procedures

We used a qualitative research design to assess the current multiple regression reporting standards. Specifically, we conducted a content analysis of two educational research journals published by the AERA: the AERJ and the *Educational Evaluation and Policy Analysis* (EEPA) journal. The Sage search engine was used to conduct a query using the key word regression, within each of the two journals. The search created a sampling frame of 590 articles in AERJ and 289 articles in EEPA. One hundred articles were then randomly selected from the sampling frames of each journal, resulting in a final sample of 200 articles. The articles were then analyzed using a qualitative analysis approach; document/content analysis (Creswell, 2003). Articles that did not contain actual multiple regression techniques were replaced. For example, some of the randomly selected articles were book reviews or made a one-line reference to the term regression but did not conduct an analysis using regression. Articles reviewed in AERJ covered 40 years ranging from 1968 to 2008 and the articles in EEPA ranged from 1979 to 2008. The difference in years occurred because the search was not restricted by years, but instead simply by the use of multiple regression and the random sampling.

Each article was then coded by two coders using the following categorizations: (a) simultaneous multiple regression, or stepwise/hierarchical regression depending on the method used to enter the independent variables, (b) whether or not corrections of Type I error rates were made for HMLR/SMLR, (c) whether authors properly identified HMLR/SMLR when used, (d) inclusion of correlation matrices, basic descriptive statistics data (e.g., mean and standard deviation), effect size statistics (e.g., R^2), standard errors, and lastly, whether *F* statistics and *t* statistics were provided.

Results

About 30% of the articles from AERJ and EEPA used HMLR/SMLR methods. Ninety percent of the articles that used HMLR/SMLR in AERJ and 95% of the articles in EEPA failed to adjust their testwise error rate. In short, of the articles that used HMLR/SMLR, only 10% of the articles in AERJ and 5% of the articles in EEPA used a procedure to ensure a reduction in Type I error rate. The remaining articles found significance when in fact, if the researchers had adjusted their testwise error rate, the results may have been different. Additionally, in light of the recommendations against the use of stepwise regression, it was noteworthy that of the 60 articles using stepwise or hierarchical regression, more than 50% of the articles were published within the last 10 years. This finding heightens the urgency of this study.

The content analysis also confirmed inconsistencies in the reporting of basic regression summary statistics, which hinders transparency and replicability. In particular, it is recommended that researchers report the following statistics when conducting regression analysis: means, standard deviations, bivariate correlations, overall *F* value, regression coefficients, R^2, and changes in R^2 (for stepwise regression methods). In Table 1, a summarization is provided listing the number of articles that included these basic data in their manuscripts.

Table 1. Percentage of Manuscripts that Included Summary Statistics

	Journal	
Descriptive Statistic	AERJ	EEPA
Means	83%	90%
Standard Deviation	77%	82%
Correlation Matrix	48%	55%
Overall *F* value	95%	95%
t statistics	80%	85%
Regression coefficients	95%	95%
R^2	95%	95%
*Change in R^2	90%	95%
Standard Errors	65%	70%

Note. *Change in R^2 is only reported for those articles that used stepwise regression

Table 2[a] Probability of Entry into a Bachelor's Program: Fall 1990 First-time Freshman (Survey Respondents) Logistic Regression

	Model 1			Model 2			Model 3		
Variable	B	SE	%	B	SE	%	B	SE	%
Immigrant origin									
Foreign born, U.S. HS	0.13	0.06	3.1*	0.30	0.09	6.8***	0.46	0.13	10.3***
Foreign born, foreign HS	-0.25	0.11	-5.8*	-0.16	0.15	-3.4	-0.07	0.16	-1.4
Race/ethnicity	--	--	--						
Black	--	--	--	-0.43	0.09	-8.5***	-0.38	0.09	-7.1***
Hispanic	--	--	--	-0.04	0.09	-0.9	-0.01	0.10	-0.2
Asian	--	--	--	0.00	0.13	0.1	-0.01	0.13	-0.2
GED	--	--	--	-0.69	0.11	-12.7***	-0.70	0.11	-12.1***
Aspirations	--	--	--	1.02	0.10	24.4***	1.02	0.10	24.2***
Gender (F = 1)	--	--	--	0.36	0.07	8.2***	0.47	0.08	10.5***
Age (minus 18)	--	--	--	-0.03	0.01	-0.6***	-0.03	0.01	-0.6
Enrolled part-time, F90	--	--	--	-0.12	0.11	-2.5	-0.10	0.11	-2.1
Supporting Children	--	--	--	-0.74	0.14	-13.4***	-0.73	0.14	-12.5***
Employment, F90	--	--	--						
Part-time	--	--	--	-0.16	0.08	-3.4**	-0.17	0.08	-3.4**
Full-time	--	--	--	-0.33	0.12	-6.6***	-0.33	0.12	-6.3***
Household income	--	--	--						
16K to 31K	--	--	--	-0.16	0.10	-3.4*	-0.15	0.10	-3.0
31K+	--	--	--	-0.15	0.10	-3.2	-0.14	0.10	-2.8
Missing income	--	--	--	-0.45	0.10	-8.8***	-0.44	0.10	-8.1***
Parent's education	--	--	--						
High school degree	--	--	--	-0.14	0.09	-3.0	-0.15	0.09	-3.0*
Some college	--	--	--	-0.11	0.10	-2.4	-0.13	0.11	-2.6
College degree	--	--	--	-0.17	0.11	-3.5	-0.19	0.11	-3.8*
Graduate/professional	--	--	--	0.18	0.13	3.9	0.16	0.13	3.4
Hybrid college	--	--	--	-1.22	0.09	-19.4***	-1.22	0.09	-18.2***
CDSEEK	--	--	--	1.28	0.10	30.7***	1.29	0.10	30.8***
Assessment tests	--	--	--						
Math	--	--	--	0.73	0.04	17.4***	0.72	0.04	16.6***
Reading	--	--	--	0.44	0.04	10.2***	0.58	0.06	13.1***
Interactions									
FB, U.S. HS*reading	--	--	--	--			-0.22	0.09	-4.2**
FB, FRGN HS*reading	--	--	--	--			-0.53	0.13	-9.5***
FB, U.S. HS*female	--	--	--	--			-0.24	0.16	-4.7
FB, FRGN HS*female	--	--	--	--			-0.56	0.26	-10.0
Constant	-0.40	0.03	40.1	-0.79	0.15	31.3	-0.91	0.16	28.8
-2 Log likelihood	7344.181			5549.253			5524.656		

Note. N = 5413 (unweighted), B = Coefficient, HS = High School, FB = Foreign Born, FRGN = Foreign, and *p < .10, **p < .05, ***p <.01. [a] Table 2 is derived from Bailey and Weininger (2002).

Adjusting Alpha Levels in Multiple Regression: An Example

Table 2 is an example from an article in the sample. In this case, the authors focused on foreign-born and native minority community college entrants at City University of New York. Stepwise logistic regression was used to predict the likelihood of entering a four-year or a two-year college program. The results show that non-native US students who immigrate to the US and graduate from a US high school are more likely than native US students to enroll in a four-year program. Additionally, those non-native US students who immigrated after high school (i.e., attended a non-US high school) are more likely to enroll in a two-year program (Bailey & Weininger, 2002).

The authors tested a total of three models. The third model is referred to as the full model and contained nearly 30 independent variables compared to only two independent variables in the initial model. Independent variables were added to the initial model until the full model was developed. All three models used the same sample. The authors report that the control variables, which were entered in model 2, had the expected influence. In particular, students who earned a GED, older students, those with jobs, those with childcare responsibilities, and those who did not aspire to a higher degree were more likely to enroll in a community college.

Interestingly, the authors report what they called a counterintuitive result concerning parental education. In the full regression model, all levels of parental education other than the highest level (i.e., attendance at graduate or professional school) exhibited negative coefficients even though only two categories were statistically significant with small effect sizes. The results suggested that if the parent had a high school degree or a college degree that those students were less likely to attend a four- year college. While these results were statistically significant, they were only marginally significant (i.e., $p < .10$). Had the researchers employed the Bonferroni correction, the testwise alpha level would have been set at .0167 and these variables would not have been statistically significant. Additionally, had the authors applied the modified Bonferroni approach proposed by Roozeboom et al. (2008), even fewer independent variables would maintain their level of significance. Roozeboom et al. posit that each time a new model is run and a decision concerning which independent variables are significant contributors is made, the Type I error rate is increased 2 to 6 times the nominal alpha level depending on the number of independent variables and their nonzero correlation with the dependent variable.

Continuing with the most basic application, given that there were three models tested in this example, employing the Bonferroni correction would result in having an adjusted alpha level of .02 (.05/3 = .017 rounded to .02) for the overall regression equation. As a result, two key interactions in the study (i.e., foreign born, US high school, reading score interaction; and foreign born, foreign high school, and female) would no longer be significant.

Regrettably, as was demonstrated, educational researchers regularly fail to adjust their testwise alpha levels when conducting HMLR/SMLR or simultaneous MLR consequently inflating their experimentwise alpha levels and Type I error rates. Zwick and Sklar (2005) is one of very few examples of authors who attempted to adjusted their testwise alpha levels. In their article published in AERJ, it was reported that "Statistical significance tests for individual predictors were conducted at an alpha level of .01 because of the large number of hypotheses being assessed" (p. 451).

Although Zwick and Sklar's (2005) article is a step in the right direction, Mundfrom et al. (2006) would suggest that they did not go far enough. The full model should evaluate each variable at the α/k level of significance, where α equals the beginning nominal alpha level and k equals the number of independent variables in the equation, which would result in an adjusted significance level of .05/30 or .002. To maintain power, the Roozeboom et al. (2008) modified Bonferroni technique could have also been applied as well.

At minimum, researchers should report the number of hypothesis tests they run (i.e., the number of models tested) so that it is clear whether the appropriate correction procedure is used. Zwick and Sklar (2005) used a common approach in education literature which is to just use .01, but if the authors had run 20 different models or hypothesis tests, the proper alpha could be .05/20 = .003. While an unusual occurrence, it is essential that researchers report the number of models or different hypothesis tests conducted.

In summary, our results demonstrated that educational researchers are not adhering to the basic standards of reporting when conducting MLR analyses (cf. Hoyt et al., 2006; Schafer, 1991). Moreover, researchers have adopted the trend to report a range of alpha levels, for example, from .05 to .01, never specifying their *a priori* alpha level. Additionally, in the discussion of their results, researchers will often report their results using a variety of significance levels; variables will be referred to as statistically significant whether at the .10, .05, or .01 level instead of maintaining a single *a priori* standard. As a result, inappropriate and possibly damaging recommendations for practice may flow from these studies.

Recommendations for Reporting Regression Results

Our recommendations for what should be included when reporting regression results are as follows: researchers should (a) describe the variables and the conceptual sets of variables (if distinct sets exist), (b) indicate if the sets are ordered, (c) describe what technique was used to adjust the alpha level when multiple models are run, (d) explicitly state the *a priori* alpha level, and (e) describe the research conclusions reached. Authors should include the following in separate tables: (a) descriptive statistics (i.e., means, standard deviations, and sample sizes), (b) correlation matrices of all continuous variables, and (c) regression results that include: the overall F ratio for each test, R^2, adjusted R^2 when comparing regression equations with different numbers of predictors and when using small sample sizes, standard error of estimate if the dependent variable has a meaningful metric, the change in R^2 and the associated significance for HMLR, regression coefficients and their associated t tests. Furthermore, we recognize page and word restrictions commonly in place in journals for authors seeking publication. In this event, we suggest authors make supplementary material available on websites or through other avenues provided through publication.

Conclusion

Results of this study indicated that researchers commonly fail to adjust alpha levels when implementing HMLR techniques. Instead, it is much more common to see authors report results using a range of alpha levels from .10 to .01, which leads to inflated Type I and experimentwise error rates. Consequently, many of the results and recommendations reported in the studies in AERJ and EEPA based on HMLR/SMLR techniques may be misleading. Additionally, about half of the articles failed to properly document research findings to ensure transparency and replication (e.g., correlation matrices). The omission of basic summary statistics not only prevents replication, but it violates the principle of transparency. Regression is a powerful statistical analysis tool when used correctly. We call for a common convention in reporting and a return to basic scientific research standards.

References

Altman, D. (2000, January). *A review of experimentwise type I error: Implications for univariate post hoc and for multivariate testing*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *35*(6), 33 – 40.

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Bailey, T., & Weininger, E. (2002). Performance, graduation, and transfer of immigrants and natives in City University of New York community colleges. *Educational Evaluation and Policy Analysis*, *24*(4), 359-77.

Clements, D., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443-494.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Courville, T. & Thompson, B. (2001). Use of structure coefficients in published multiple regressions articles: B is not enough. *Educational and Psychological Measurement*, *61*(2), 229-248.

Creswell, J.W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Daniel, L. (2001). *Changes in the APA "Publication Manual": How the new fifth edition will affect research reporting in the social sciences*. Retrieved from ERIC database.

Desimone, L, Smith, T., Baker, D., & Uano, K. (2005). Assessing barriers to the reform of U.S. mathematics instruction from an international perspective. *American Educational Research Journal*, *42*(3), 501-535.

Elmore, P. B., & Woehlke, P. L. (1996, April). *Research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1995.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences* (7th ed.). Belmont, CA: Wadsworth Publishing.

Hoyt, W. T., Leierer, S., & Millington, M. J. (2006). Analysis and interpretation of findings using multiple linear regression techniques. *Rehabilitation Counseling Bulletin*, *49*(4), 223-233.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499-534.

Keppel, G., & Wickens, T.D. (2004). *Design and analysis: A researcher's handbook* (4th Ed.). Upper Saddle River, NJ: Prentice Hall.

Love, G. (1988, November). *Understanding experimentwise error probability*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Louisville, KY.

McInerney, D.M., Roche, L., McInerney, V., & Marsh, H. W. (1997). Cultural perspectives on school motivation: The relevance and application of goal theory. *American Educational Research Journal*, *34*(1), 207-236.

Mundfrom, D. J., Perrett, J. J., Schaffer, J., Piccone, A., Roozeboom, M. (2006). Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints*, *32*(1), 1 – 6.

Olejnik, S., Li, J., Supattathum, S., & Huberty, C.J. (1997). Multiple testing and statistical power with modified Bonferroni procedures. *Journal of Educational and Behavioral Statistics*, *22*, 389-406.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F.J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*(2), 365-397.

Roozeboom, M., Mundfrom, D. J., & Perrett, J. (2008, April). A modified Bonferroni procedure for multiple tests. Paper presented at the annual meeting of the American Educational Research Association, New York.

Salkind, N. J. (2007). *Statistics for people who think they hate statistics* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Schafer, W. (1991). Reporting hierarchical regression results. *Measurement and Evaluation in Counseling and Development*, *24*(3), 98-100. (ERIC Document Reproduction Service No. EJ438926) Retrieved from ERIC database.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Sidak, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626-633.

Stern, D., Dayton, C., Paik, I. W., & Weisberg, A. (1989). Benefits and costs of dropout prevention in a high school program combining academic and vocational education: Third-year results from replications of the California peninsula academies. *Educational Evaluation and Policy Analysis*, *11*(4), 405-416.

Thompson, B. (1995). Stepwise regression and stepwise discriminate analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, *55*, 525-534.

Thompson, B. (2007, October). *Standards in conducting and publishing research in education*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, St. Louis, MO.

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, *38*(5), 343-352.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, *42*(3), 439-464.

Send correspondence to: Winona Burt Vesey
University of Houston, Clear Lake
Email: Vesey@UHCL.edu