Suppressor Variables: The Difference between 'Is' versus 'Acting As'

Larry Ludlow
Kelsey Klein
Boston College

Journal of Statistics Education Volume 22, Number 2 (2014),
www.amstat.org/publications/jse/v22n2/ludlow.pdf

Key Words: Suppressors; Suppression effects; Multicollinearity; Course evaluations; Student ratings of instruction

Abstract

Correlated predictors in regression models are a fact of life in applied social science research. The extent to which they are correlated will influence the estimates and statistics associated with the other variables they are modeled along with. These effects, for example, may include enhanced regression coefficients for the other variables—a situation that may suggest the presence of a suppressor variable. This paper examines the history, definitions, and design implications and interpretations when variables are tested *as* suppressors versus when variables are found that *act as* suppressors. Longitudinal course evaluation data from a single study illustrate three different approaches to studying potential suppressors and the different results and interpretations they lead to.

1. Introduction

The presence of correlated predictors in ordinary least squares (OLS) multiple regression models is routinely encountered in applied social science research. More formally known as multicollinearity, or simply collinearity, the introduction of a second predictor (X_2) into a model along with the original predictor of interest (X_1) will change the regression coefficient, standard error, significance test value, and p-value associated with X_1 (when the two predictors are correlated). In this situation the regression coefficient for X_1 may be diminished or enhanced and even reversed in sign. As additional correlated predictors are added to the model, all previous estimates associated with X_1 will again change (Pedhazur 1997). In our teaching experience these statistical changes are usually not difficult conceptually or technically to explain. The problem for researchers and students, however, lies in how the changes in real, and often messy, data are understood and interpreted.

As manuscript and dissertation reviewers we have observed that depending on how the regression model is constructed (i.e. all predictors entered simultaneously as a block or entered in a systematic one-at-a-time hierarchical forced-entry manner) and how experienced and well trained the researcher is, the changes in coefficients may be: a) unknown because they are not noticed, b) ignored because they are not understood, or c) described as the consequence of a variable that *is* or *acts as* a mediator or a suppressor (depending on the direction of the change) in order to suggest some plausible mechanism is at work when the results were, in fact, unexpected. That mechanism, however, is not the same for mediator and suppressor variables.

In fact, the phrase "*is* a mediator" is properly reserved for the deliberate, i.e. experimental, introduction of an intervention variable that accounts for some of the covariance between X_1 and the criterion (Baron and Kenny 1986; Dearing and Hamilton 2006; MacKinnon, Krull, and Lockwood 2000). The mechanism underlying suppressor variables, however, is solely statistical—no causal intervention is assumed to produce suppressor effects as is the case with mediators. The use of *is*, in the present paper, will therefore reflect the intentional introduction into the model of a variable hypothesized to strengthen the relationship between the variable of interest and the criterion, rather than an experimental intervention designed to produce that effect. Hence, the distinction between *acts as* a suppressor and *is* a suppressor is limited to differentiating between a variable that demonstrates a statistical effect potentially devoid of substantive interpretation versus a variable's effect supported by theory, respectively.

This distinction is important because one of the points of the present exercise is that a long-standing refrain in teaching multiple regression is to encourage model building with uncorrelated predictors (Horst 1941, p. 435), each of which is correlated with the criterion (Kerlinger and Pedhazur 1973, p. 46). What makes a suppressor variable so interesting is that it is typically correlated with another predictor but not the criterion (this and other definitions are presented later)—and this situation, as we shall see, is recommended as a good design strategy. This strategy, however, is not intuitively obvious. Unlike mediator variables whose effects are often presented as some form of intervention under the investigator's control, suppressors are not typically theorized and discussed in the introduction sections of articles. In fact, when working with our clients and colleagues, suppressors have usually been thought about *post hoc* when a variable was added to a model, and a previously estimated coefficient became unexpectedly stronger. Such an observation typically produced confusion followed by imaginative substantive interpretations offered as pseudo-explanations of what "caused" the changes. Not once, in forty years of statistical work, has the first author been approached to work on an *a priori* suppressor variable design even though, as we intend to illustrate, such designs may be reasonable, justifiable, and powerful.

Standard education and psychology textbooks do not typically delve deeply into the origins and history of most of the commonplace statistical procedures covered in applied social science statistics courses (e.g., Hinkle, Wiersma, and Jurs 2003; Field 2013). This is understandable to a certain extent, but it is also unfortunate because the original conceptualization of a procedure can provide guidance about intended uses and limitations that complement, yet cannot be captured by, the modern emphasis on matrix algebra expressions of the mechanics of the procedure or the simple execution of statistical software. It is this purpose, i.e. understanding the origins and purposes of suppressor design and analysis, which this paper addresses.

In the following sections we present the origin of suppressor variables, various definitions of what they are and how they may be identified, how they have been treated in research designs, how they are related to other so-called third variable effects, an application with real data that illustrates results and interpretations depending on which of three proposed suppressor design or suppression analysis strategies are taken, and recommendations for teaching and research.

1.1 History of suppressor variables and effects

The following section is not an exhaustive review of the suppressor variable or suppression effects literature. Friedman and Wall (2005), in contrast, provide an extensive technical review that incorporates a comparison of terminology and useful graphical representations of correlational, including suppression, relationships. Rather, our review highlights some of the main conceptual issues while illustrating the challenges in teaching this often confusing topic. Along these lines, it is important to clarify that in the present paper *validity* refers to the correlation between a *predictor* and *criterion*, the use of *independent* and *dependent* variables is reserved for experimental designs, and shared variance between predictors is *partialled* rather than “held constant” or “controlled”.

The original term for suppressor variables was “clearing variates,” coined by Mendershausen (1939) to define “a useful determining variate without causal connection with the dependent variate; its rôle in the set consists of clearing another determining (observational) variate of the effect of a disturbing basis variate” (p. 99). In essence this means that when a predictor variable (X_2) is not correlated with a criterion (Y) but is correlated with another predictor (X_1) and is entered into the model after X_1, X_2 will remove extraneous variation in X_1. This partialling out of extraneous variation would presumably then clarify or “purify” and thereby strengthen the relationship between X_1 and the criterion.

Although not pointed out by Mendershausen, the concept of a variable consisting of disturbing variance (his term) and non-disturbing variance (our term) is consistent with the True Score Theory (or Classical Test Theory) definition of an observed variable’s variance made up of common, specific and error variance wherein common and specific variance constitute true score variance (Gulliksen 1950; Spearman 1904). Hence, removing error variance should yield a clearer understanding of an observed variable’s characteristics and “make the original predictor more valid” (Lord and Novick 1968, p. 271-272). Their description, implying that the partial correlation between a predictor and criterion will be greater than the zero-order correlation, subsequently came to be called “Lord and Novick’s intuitive explanation” of suppressor effects (Tzelgov and Stern 1978, p. 331), and this partitioning of variance is what Smith, Ager and Williams (1992, p. 21) later refer to as “valid variance” and error.

Horst (1941) expanded the concept by suggesting that a “suppressor variable” is a predictor that “should be relatively independent of the criterion but the major part of its variance should be associated with that variance of the prediction system which is independent of the criterion" (p. 141). Furthermore, he suggested "it would be extremely valuable to have a much more general and exhaustive analysis of the problem than is given in this study" (Horst 1941, p. 141). This suggestion immediately produced algebraic specifications of suppressor effects (McNemar 1945; Meehl 1945; Wherry 1946). McNemar's subsequent work on "suppressants" turned Horst's definition into an explicit recommendation for variable selection. He stated "it is possible to increase prediction by utilizing a variable which shows no, or low, correlation with the criterion, *provided* it correlates well with a variable which does correlate with the criterion" (McNemar 1949, p.163). Wiggins (1973), too, extended Horst's definition by suggesting that one rationale for the selection of suppressor variables is that they "should lead to a *practical* increment in predictive validity" (p. 32). Conger (1974), also referring to predictive validity, later refers to Horst's definition as "traditional" while Tzelgov and Stern (1978) refer to Horst's definition as "classic".

These early definitions are important because they address two points that subsequent authors have tried more or less successfully to expand upon. That is, the effect of Mendershausen's clearing variate is to increase the strength of the relationship between the outcome and the predictor of interest while the intention of Horst's and McNemar's suggestions are to investigate and test possible candidates that would contribute to this effect. One tells us what happens in the presence of a suppressor while the others tell us how to look for them.

There followed a proliferation of research to more precisely define the presence and characteristics of suppressors. For example, Lubin (1957) introduced "negative suppressors," and Darlington (1968) discussed "negative weights." Conger (1974) revised their definitions and proposed an additional situation he called "reciprocal suppression." Cohen and Cohen (1975) classified the existing definitions into four categories including cooperative suppression, redundancy, net suppression, and classical suppression (p. 84-91). Tzelgov and Stern (1978) subsequently extended Conger's definition and identified three different types of negative suppressors, and Velicer (1978) added a new definition in terms of semipartial correlations—a definition which we employ later.

McFatter (1979) then pointed out that these various patterns of relationships "to *technically* define and diagnose 'suppression' can be generated by a large number of causal structures [e.g. structural equation models] for which the notion of suppression as the measuring, removing, or subtracting of irrelevant variance is inappropriate and misleading" (p. 123). He introduced the term "enhancers" to refer to these technical occurrences of suppression and reserved the term suppressor for the particular two-factor model discussed by Conger (1974). Pedhazur (1982) simplified the conversation by suggesting that a suppressor variable is involved "when a partial correlation is larger than its respective zero-order correlation" (p. 104)—a position supported by Kendall and Stuart (1973, p. 331) in their earlier discussion of a "masking variable." Holling (1983), extending both Conger's and Velicer's definitions, suggested conditions for suppression within the General Linear Model. Tzelgov and Henik (1985) then proposed a definition in terms of regression weights that was consistent with Holling's and equivalent to Conger's.

Currie and Korabinski (1984), in addressing common errors in the context of bivariate regression, introduced "enhancement," a term they distinguished from suppression yet is easily confused with McFatter's enhancers and, furthermore, is referred to as suppression by Sharpe and Roberts (1997) (see, Friedman and Wall 2005, on Sharpe and Roberts 1997). Hamilton (1987), in turn, suggested the term "synergism" instead of enhancement and Shieh (2001) subsequently combined these two ideas to form "enhancement-synergism."

Pedhazur (1997) contributed an important insight by pointing out that McFatter's enhancers "may be deemed *suppressors* by researchers whose work is devoid of a theoretical framework" (p. 188). Pedhazur's point requires elaboration. Pedhazur was a student of Fred Kerlinger (see Pedhazur 1992). Early in his career, Kerlinger (1964) made the useful distinction between explanation and prediction. For Kerlinger, explanation of phenomena was the aim of science. Prediction, in contrast, did not necessarily require explanation. These distinctions then led to considerations of how statistical models were constructed and interpreted. For Kerlinger and Pedhazur (1973) and then Pedhazur (1982, 1997), statistical models, particularly regression models, were either theory-based (i.e., explanation) or built to maximize R^2 (i.e., prediction). The distinction between explanation and prediction is useful because it offers a way to distinguish between the identification and interpretation of suppressor variables versus suppression "effects" or "situations" (Tzelgov and Henik 1991).

Suppression effects and situations attracted attention around the turn of the century as numerous articles proposed multiple criteria to determine their presence (e.g., Pandey and Elliot 2010; Voyer 1996; Walker 2003). Voyer (1996), for example, suggested three related criteria for identifying a suppression effect (p. 566). In the context of explanation versus prediction, suppression effects or situations do not necessarily require the identification and testing of a particular variable; therefore, they do not necessarily require a theoretical framework. Interestingly, although Voyer (1996) uses his criteria to determine "whether mathematical achievement *acts as* a suppressor variable on gender differences in spatial performance" (italics added), his three experiments and psychological interpretations of the hypothesized and observed suppression effects clearly show he was testing whether mathematical achievement *is* a suppressor, not simply *acting as* a suppressor.

In addition to these efforts to examine situations under which suppression may occur, formal specifications were offered. Hamilton (1987), for example, provided the necessary and sufficient condition for Kendall and Stuart's (1973) "masking variable." Schey (1993), extending Hamilton (1987, 1988), expressed the condition for suppression in terms of the geometry of angles and cosines. Sharpe and Roberts (1997), in contrast, proposed a necessary and sufficient condition based directly on correlation coefficients—an advantage in that "cases of suppression can be identified directly from the correlation matrix" (p. 47).

Meanwhile, significant research coming from educational psychology, sociology, and statistics extended the concept of suppression beyond the bivariate situation. Smith, Ager, and Williams (1992), Maassen and Bakker (2001), Lynn (2003), and Shieh (2006), for example, discussed suppression with more than two predictors, argued for the equivalency of different definitions and offered extensions to other statistical methods.

Finally, during the first decade of the present century, another body of research appeared comparing suppressors to other so-called third variables such as mediators and confounders

(Logan, Schatschneider, and Wagner 2009; MacKinnon et al. 2000; Stewart, Reihman, Lonky, and Pagano 2012; Tu, Gunnell, and Gilthorpe 2008). While mediators and confounders represent different situations, both typically assume "that statistical adjustment for a third variable will reduce the magnitude of the relationship between the" criterion and the predictor (MacKinnon et al. 2000, p. 174). However, it is possible that the third variable "could increase the magnitude of the relationship"; this "would indicate suppression" (MacKinnon et al. 2000, p. 174).

1.1.1 Three distinct strategies

A teacher of applied regression could be excused at this point for being uncertain about what and how to teach and illustrate different aspects of suppression without hopelessly confusing students. For example, in our experience Table 1 of MacKinnon et al. (2000), in which 18 possible outcomes of mediation, confounding, suppression (also called "inconsistent mediation" or "negative confounding") and chance are mapped out, is useful in an academic sense but not useful as a pedagogical device. Likewise, the simulated hypothetical data and examples in Tu et al. (2008) provide plausible situations in which suppressor effects may occur, but they lack the authenticity of empirical hypothesis confirmation or surprise findings.

Our review of the suppressor variable literature suggests three distinct strategies that have been used when testing a suppressor, or exploring or claiming the presence of a suppression effect: *theory-based hypothesis testing*, *maximizing predictive validity variance*, and *post-hoc determination*, respectively. The next section presents the statistical details of the third variable, or bivariate, regression model that underlies these three different strategies.

1.2 Statistical detail

The statistical equivalence of various third variable regression models is typically established by relying on correlation notation (e.g., Tu et al. 2008). We find this approach, when teaching regression, to be more obfuscating than clarifying. However, if one prefers correlation notation, we recommend Sharpe and Roberts (1997) and Friedman and Wall (2005) for their presentation on the identification of suppressor relationships through correlations. We prefer sums of squares notation since it shows more clearly at a basic calculation level how the variances and covariances combine to form the various regression estimates. The notation is too unwieldy for more than two predictors but it is ideal for showing what happens when a second predictor is added to an OLS regression model.

The population regression model is:

$$\mathrm{Y_i} = \beta_0 + \beta_1 \mathrm{X_{1i}} + \beta_2 \mathrm{X_{2i}} + \cdots + \beta_k \mathrm{X_{ki}} + \epsilon_i,$$

where $\mathrm{X_1}, \ldots, \mathrm{X_k}$ are the predictor variables, $\beta_1 \ldots, \beta_k$ are the population regression coefficients, k is the number of predictors, i denotes the $\mathrm{i^{th}}$ observation, and ϵ_i is a random error. The corresponding sample regression model is:

$$\mathrm{Y_i} = \mathrm{a} + \mathrm{b_1 X_{1i}} + \mathrm{b_2 X_{2i}} + \cdots + \mathrm{b_k X_{ki}} + \mathrm{e_i}.$$

A zero-order correlation between X_1 and X_2 ($\mathrm{r_{12}}$) represents *collinearity*. Ideally, for the purpose of interpreting the partitioning of unique variance in Y, $\mathrm{r_{12}} = 0$ and $\mathrm{R^2_{y12}} = \mathrm{r^2_{y1}} + \mathrm{r^2_{y2}}$ as seen in Figure 1, Diagram 1.

Figure 1. Venn diagrams of two possible predictor relationships

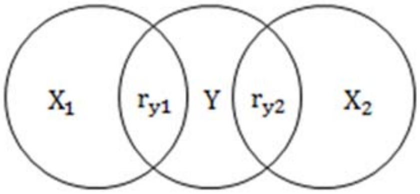

Diagram 1 Zero predictor collinearity

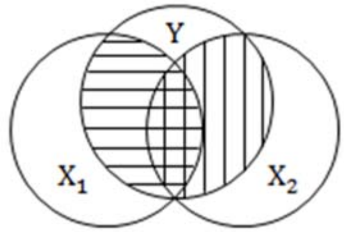

Diagram 2 Non-zero predictor collinearity

In most, if not all, social science applications the reality is that $r_{12} \neq 0$ and $R^2_{y12} = \frac{r^2_{y1} + r^2_{y2} - 2r_{y1}r_{y2}r_{12}}{1 - r^2_{12}}$. An apparently simple $r_{12} \neq 0$ situation is represented in Diagram 2 where the horizontally lined area (including the cross-hatched area) represents r_{y1}, the vertically lined area (including the cross-hatched area) represents r_{y2} and the cross-hatched area itself represents r_{12}. As additional correlated predictors are added, the overall R^2 tends to increase but the unique contribution of each predictor usually becomes conceptually and statistically difficult to disentangle.

Although these Venn diagrams, or ballantines (Cohen and Cohen 1975, p. 80), have a comfortable appeal in their simplicity, they are widely criticized as overly simplistic, misleading, and wrong in many situations. McClendon (1994), for example, uses them to depict certain correlational relationships and describes when these diagrams cannot appropriately depict situations such as suppression (p. 114). Others have offered alternative graphs and charts for depicting correlational relationships that include suppression (Friedman and Wall 2005; Schey 1993; Sharpe and Roberts 1997; Shieh 2001, 2006). The geometric illustrations in Figure 5 by Currie and Korabinski (1984) are particularly creative and insightful.

Nevertheless, the different sources of covariation represented in Figure 1 are useful to bear in mind when we consider the formulas for the coefficients of the linear regression model. Using deviation from the mean notation ($x_{ki} = X_{ki} - \overline{X}_k$), when k = 1 the regression equation is

Equation 1 $$Y' = a + b_{Y1}X_1,$$

the regression coefficient for predictor X_1 (where Σ is across i=1,N) is

Equation 2 $$b_{Y1} = \frac{\Sigma x_1 y}{\Sigma x_1^2},$$

and the standard error of the regression coefficient for predictor X_1 is

Equation 3 $$se_{b_1} = \sqrt{\frac{\Sigma(Y - Y')^2/N-k-1}{\Sigma x_1^2}}.$$

When a second predictor is added and k = 2, the expression takes into account the sum of squares of X_2 and the sum of cross products of X_2 with X_1 and Y. Hence, the regression equation is

Equation 4 $$Y' = a + b_{Y1.2}X_1 + b_{Y2.1}X_2,$$

the regression coefficient for predictor X_1 is now

Equation 5 $$b_{Y1.2} = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2},$$

and the standard error of the regression coefficient for predictor X_1 is now

Equation 6 $$se_{b_{1.2}} = \sqrt{\frac{\sum(Y-Y')^2/N-k-1}{\sum x_1^2(1-R_{12}^2)}}\,.$$

When k = 2 and $\sum x_1 x_2 = 0$, then collinearity is zero, and the equations for b_{Y1} and $b_{Y1.2}$ are equal. However, when $\sum x_1 x_2 \neq 0$ then collinearity is non-zero and $b_{Y1} \neq b_{Y1.2}$.

By the definitions presented earlier, if "the strength of the relationship between the predictor and the outcome is reduced by adding the [second predictor]" (Field 2013, p 408), or in our terms, $|b_{Y1.2}| < |b_{Y1}|$, then X_2 *is* (or, conversely, *acts as*) a mediator. "[When] the original relationship between two variables increases in magnitude when a third variable is adjusted for in an analysis" (MacKinnon 2008, p. 7), or in our terms, $|b_{Y1.2}| > |b_{Y1}|$, then X_2 *is* (or, conversely, *acts as*) a suppressor variable.

The point of this section was to establish that the estimate for $b_{Y1.2}$ is determined the same way regardless of what kind of third variable X_2 *is* or *acts as*. Furthermore, depending on the cross product patterns for $\sum x_1 y$, $\sum x_1 x_2$ and $\sum x_2 y$, $b_{Y1.2}$ may change sign or stay the same sign but increase or decrease relative to b_{Y1}. This means that the purpose for conducting the analysis must be clear at the outset of the design or else the risk of misinterpretation of the findings is real and non-trivial.

1.2.1 Test of significance

Since the third variable estimation procedure is the same regardless of what the second predictor is called, the well-established tests of Sobel (1982) and Freedman and Schatzkin (1992), from MacKinnon, Lockwood, Hoffman, West, and Sheets (2002), for the statistical significance of a mediator's effect also hold for testing a suppressor's effect (MacKinnon et al. 2000, p. 176). Furthermore, the Sobel (1982) and Freedman and Schatzkin (F-S) (1992) tests themselves are equivalent (MacKinnon, Warsi, and Dwyer 1995).

We recognize that even though the Sobel and F-S tests are frequently employed with the influential "causal steps strategy" (Preacher and Hayes 2008, p. 880) for testing mediation proposed by Kenny and his colleagues (Baron and Kenny 1986; Judd and Kenny 1981; Kenny, Kashy and Bolger 1998), the tests have their critics, flaws and replacement strategies when the OLS regression assumptions are violated. For example, after providing an excellent review of the four-step Kenny framework, Shrout and Bolger (2002) illustrate bootstrap methodology introduced for mediator models by Bollen and Stine (1990) and applied to small samples by Efron and Tibshirani (1993). MacKinnon, Lockwood and Williams (2004) address mediator effect confidence interval estimation by testing the power and Type I error rates of Monte Carlo, jackknife and various bootstrap resampling methods. Iacobucci (2008) discusses directed acyclic graphs and structural equation model approaches. And MacKinnon (2008), in covering the work on mediation from the 3rd century BC to the present, conveniently provides software code and data for everything from single mediator problems to longitudinal and multilevel designs and statistical tests.

In summary, mediator, or indirect effect, tests exist in many forms for different purposes. Our pedagogical purpose in using the F-S test is to demonstrate the essential nature of all these tests, namely, the difference between b_{Y1} and $b_{Y1.2}$. Specifically, the F-S test is based on the difference between the adjusted (first-order partial) and unadjusted (zero-order) regression coefficients for X_1. Modifying their notation to suit a suppressor variable problem, the hypotheses are H_0: $\tau - \tau' = 0$ and H_1: $|\tau| - |\tau'| < 0$. The test statistic is

Equation 7
$$t_{N-2} = \frac{\tau-\tau'}{\hat{\sigma}_{F-S}}$$

where $\tau = b_{Y1}$, $\tau' = b_{Y1.2}$, $\hat{\sigma}_{F-S} = \sqrt{\hat{\sigma}_{\tau}^2 + \hat{\sigma}_{\tau'}^2 - 2\hat{\sigma}_{\tau}\hat{\sigma}_{\tau'}\sqrt{1-\hat{\rho}_{X_1X_2}^2}}$, $\hat{\sigma}_{\tau}$ is the standard error of b_{Y1}, $\hat{\sigma}_{\tau'}$ is the standard error of $b_{Y1.2}$, and $\hat{\rho}_{X_1X_2}$ is the correlation between the predictor X_1 and the hypothesized suppressor X_2.

In the next two sections we describe our data and then offer five examples to illustrate how the differences in the three suppressor design and analysis strategies inform how hypotheses are framed, variables are selected, models are constructed and tested, and results are interpreted. The results are summarized in Table 1 and are presented at this point for convenient reference.

Table 1. Statistical Summary of the Analyses*

Strategy #1: Hypothesis testing	
Example 1	*predicted principles* = 44.230 + .404*(*outclass*) p < .001 R^2 = .242, p < .001 *predicted principles* = 43.582 + .398*(*outclass*) + .386*(*resactive*) p < .001 p = .545 R^2 = .245, p < .001
Example 2	*predicted principles* = 33.096 + .396*(*rgattend*) p < .001 R^2 = .153, p < .001

	predicted principles = 23.813 + .460*(*rgattend*) + 1.740*(*resactive*) $p < .001$ $p < .05$ $R^2 = .205$, $p < .001$ $\hat{\rho}_{x_1 x_2} = -.265$ F-S test statistic: $t = -2.585$, $p<.05$
Example 3	*predicted principles* = 53.506 + .180*(*time*) $p < .005$ $R^2 = .079$, $p < .01$ *predicted principles* = 49.974 + .193*(*time*) + 1.139*(*resactive*) $p < .005$ $p = .103$ $R^2 = .103$, $p < .01$ $\hat{\rho}_{x_1 x_2} = -.055$ F-S test statistic: $t = -3.938$, $p<.05$
Strategy #2: Maximizing explained validity variance	
Example 4	*predicted principles* = 53.506 + .180*(*time*) $r^2_{y1} = .079$ $R^2 = .079$, $p < .01$ *predicted principles* = 31.391 + .291*(*time*) + 12.315*(*interact*) $r^2_{y(1.2)} = .132$ $R^2 = .133$, $p < .01$
Strategy #3: *post-hoc*	
Example 5	*predicted excell* = –13.219 – .016*(*rgattend*) + .799*(*principles*) + .114*(*skills*) + .162*(*outclass*) – .124*(*time*) $R^2 = .675$, $p < .001$

*Note: the statistical details reported within each of the three sections above differ depending on the specific nature of the suppressor or suppression effect strategy employed. Details are provided later in the text addressing each strategy.

2. Method

2.1 Participants

The data are end-of-semester course evaluation student ratings of instructions (SRIs) from a single university instructor's courses. The data set consists of 110 records summarizing the SRIs for the instructor's graduate courses in applied statistics and research methods from fall 1984 through spring 2011. There were 2,234 students enrolled in these courses.

2.2 Data description

The course evaluations consisted of between twenty-two and twenty-nine questions but the specific questions being analyzed remained the same over the entire 28-year time frame. The data may be summarized into four categories: 1) student-level perceptions (e.g., time spent on the course compared to other courses, the extent to which the student understood principles and concepts), 2) administrative characteristics (e.g., year taught, class size, level of students), 3) instructor-specific variables (e.g., tenure status, rank, marital status), and 4) summative instructor evaluation ratings (percent of students in the course who marked excellent, very good, good, acceptable, and poor).

Student-level perceptions are recorded as the percent of students in each class who strongly agreed with the statement, except for time spent on the course, which is recorded as the percent that indicated that they spent more time on the course than on other courses. The instructor variables are coded categorical variables. For example, tenure status is coded 1 pre-tenure and 2 post-tenure. Administrative characteristics consist of a combination of continuous and categorical variables. There are a total of 35 variables associated with each class record. When used in combination, these administrative, student, and instructor variables provide a complex picture of the teaching and learning environment within which a class was taught, and they may be used to examine, and test, a variety of pedagogically meaningful relationships (Burns and Ludlow 2005; Chapman and Ludlow 2010; Ludlow 1996, 2002, 2005; Ludlow and Alvarez-Salvat 2001). Although the unit of analysis is the class, interpretations of results are deductions (based on the previously cited research, personal teaching experience, and conversations with students) about instructor behaviors, classroom learning conditions, and what students may have been considering when they completed the evaluations.

3. Analysis strategies

3.1 Strategy #1 (Hypothesis testing)

An *a priori* hypothesis formulation and testing approach to designing and executing a research agenda is often preferred, when possible, because of its powerful theoretical implications (Kerlinger 1964). This is because hypotheses building upon previous research offer an explicit test of, and contribution to, the literature under study. To come up with *a priori* hypotheses, however, one has to have mastered the research problem and the characteristics of the data. This familiarity with the problem and data then provides the opportunity for formulating specific expected outcomes. This sequential theoretical specification is familiar to every reader of this journal. *A priori* hypotheses about suppressor variables, although rare, follow the same train of thought (e.g., Cronbach 1950; Dicken 1963; Voyer 1996; Voyer and Sullivan 2003).

The criterion of interest for the three examples under Strategy #1 is 'understood *principles* and concepts' (the percent of students in a class who strongly agreed that they understood principles and concepts). *Principles* is a good indicator of student learning in a course, and it is useful to know what instructional and classroom environment variables are associated with an increase in this particular measure of student learning (Chapman and Ludlow 2010).

Three different predictors were selected for the following examples based on their roles in the previously cited course evaluation research: *outclass* (the percent who strongly agreed the professor was available outside of class), *rgattend* (the percent who strongly agreed that regular class attendance was necessary), and *time* (the percent of time spent on the course more than other courses). These three predictors, in their original form as survey items, are routinely included on course evaluations because it is assumed by many faculty and administrators that availability outside of class, the necessity for students to attend class, and the time that students spend on classes are all reasonable indicators of course rigor, quality and student learning. In the previous research, each of these predictors had a positive, statistically significant relationship with *principles*.

The new hypotheses for the present work concern the presence of suppressors. No previous research with these data had considered *a priori* whether any of the variables might be a suppressor. But, given our research experience with these data, the statistics lectures where these data are used to great effect to illustrate various conceptual and technical points, and the fact that faculty do express concerns about the struggle to balance research productivity with pedagogical effectiveness (Fox 1992; Ramsden and Moses 1992), we hypothesized that the number of discrete research-related activities this instructor engaged in during a semester (*resactive*) – calculated as the sum of manuscripts in preparation, workshops and seminars conducted, and conference papers and invited addresses presented – would be a suppressor for each of the three predictors in separate OLS regression models.

The *resactive* variable was constructed and added to the dataset because the instructor was well aware that in recent years his research agenda took up more of his work load, his research related travel was increasing, and his reliance on graduate students to cover lectures and provide student support was increasing—and he was concerned about the negative effects this situation might be having on his teaching. *Resactive* as a simple count, however, is only a crude proxy for the extent to which research activities can reach a level at which they become a burden and a source of stress.

The suppressor hypothesis is based on the following argument: given that the three predictors (*outclass*, *rgattend*, and *time*) are pedagogically meaningful, are to a certain extent under the control of the instructor, and are statistically significant predictors of at least one aspect of student learning (*principles*), the "true" relationship between these three predictors and *principles* may be masked by the extent to which an instructor is research active (*resactive*). This is because some of the variance in these predictors may be a function of *resactive*, i.e. diminished ratings on *outclass*, *rgattend*, and *time* due to instructor absence or unavailability. The explicit nature of the hypothesized suppressor relationship between *resactive* and each predictor is discussed and investigated in "Example 1" through "Example 3".

For Example 1, we theorized that when an instructor is involved in increasing numbers of research activities (*resactive*) some students may be less likely to strongly agree that their professor is available for help outside of class (*outclass*). Hence, if *resactive* = 0, then the student ratings provided for *outclass* will be unaffected by *resactive* but when *resactive* > 0, then the mean class rating for *outclass* may be diminished resulting in a negative relationship between *resactive* (the suppressor) and *outclass* (the predictor).

What about the relation between *resactive* and *principles* (the criterion)? Or, more broadly, how do the instructor and students play a mutually interactive role in influencing ratings of course variables and their relationships as *resactive* increases? On the one hand, some readers will argue that their research activities enhance their teaching and the learning opportunities of their students. In this scenario the instructor is fired-up, energized and enthusiastic when entering the classroom; excited while sharing points about research procedures, analyses, and findings; and, overall, fully prepared and engaged in the positive dynamics of research and teaching (Ludlow et al. 2014). Classes taught during these exciting times will likely reflect the effect of research activities through higher ratings.

On the other hand, some will argue, perhaps admit that at some point on the research activity continuum teaching begins to suffer, and student learning is adversely affected. This scenario is one of stress, annoyance and distraction; hasty, sloppy and fragmented course materials; and choppy presentations and cursory interactions with students. Classes taught under these conditions will likely reflect the distractions through lower ratings. Over a career a research active instructor may experience both situations as research activities change, in which case, the linear relationship between *resactive* and *principles* should be zero—although a quadratic relationship might exist. This tension, and aim for the right personal balance between research activity/productivity and teaching quality, exists in every academic research unit and is often an explicit component of an instructor's annual review.

Obviously, when *resactive* = 0, then *principles* cannot be rated lower by students as a consequence of an instructor's research activities that did not occur, but what about student behavior and learning when *resactive* > 0? Regardless of how effective an instructor is at integrating research and teaching, it seems plausible that when an instructor becomes unavailable because research activities have increased to the point of hindering course preparation and teaching, some students may seek out other sources of help in order to better understand course principles and concepts (*principles*) and may, therefore, still rate *principles* relatively high. Other students, however, may not be able to do anything extra in terms of seeking help in understanding *principles* and be "harmed" by the instructor's active research schedule and, consequently, rate *principles* relatively low. Hence, higher *principles* ratings attributable to extra non-instructor help when *resactive* is greatest are offset by lower *principles* ratings due to a lack of help when *resactive* is greatest. This means the relationship between *resactive* and *principles* should be zero—a null hypothesis we expect to retain.

In summary: if the relationship between (a) *outclass* (predictor) and *principles* (criterion) is positive, (b) *resactive* (suppressor) and *outclass* (predictor) is negative, and (c) *resactive* (suppressor) and *principles* (criterion) is zero, then the effect of *resactive* in partialling out "irrelevant" variance in *outclass* should be to enhance the relationship between *outclass* and *principles*.

To see how the suppression of "irrelevant" variance in *outclass* (X_1) would work, consider Equation 5. If *resactive* (X_2) and *principles* (Y) share no covariance, then $\sum x_2 y = 0$; therefore $(\sum x_1 x_2)(\sum x_2 y) = 0$. This means the numerator portion representing the cross product sum of squares between *outclass* and *principles* and the sum of squares of the suppressor $(\sum x_1 y)(\sum x_2^2)$ is unchanged. But more importantly, the denominator decreases because the product of the total predictor and suppressor sum of squares available $(\sum x_1^2)(\sum x_2^2)$ is reduced by the shared predictor and suppressor cross products $(\sum x_1 x_2)^2$. Hence, $|\mathrm{b}_{\mathrm{Y1.2}}| > |\mathrm{b}_{\mathrm{Y1}}|$.

For Example 2, we theorized that if an instructor is away on many research activities (*resactive*) then some students may be less likely to think that regular attendance (*rgattend*) is necessary—a negative relationship. Similar to the thinking outlined above, some students may adjust to the instructor's absence and develop their understanding of *principles* through other means than in-class, while others may be "harmed" by the absence. The relationship between *resactive* and *principles*, as in Example 1, should be zero, and the relationship between *rgattend* and *principles* should be strengthened.

For Example 3, we theorized that if an instructor's attention to teaching is distracted by many research activities (*resactive*) then it may be necessary for some students to spend more time on the course (*time*) than on other courses—a positive relationship that actually represents an undesirable learning situation. Here again some students may spend more time on their own developing their understanding of *principles*, while others may be limited in the additional time they have and then suffer "harm" because of the instructor's attention to research. Consistent with Examples 1 and 2, we expect a zero relationship between *resactive* and *principles*, and a strengthened relationship between *time* and *principles.*

To test each set of hypotheses, the criterion *principles* was first regressed on each separate predictor. If a statistically significant relationship was found, then the hypothesized suppressor *resactive* was added to the model. If the addition of *resactive* increased the coefficient of the predictor, the correlation between the predictor and *resactive* was tested as a check on how the results best fit with the various suppressor definitions offered previously. Then the F-S test was applied to determine the statistical significance of the effect of *resactive* upon the predictor's relationship with the criterion. Since the correlation between the suppressor *resactive* and the criterion *principles* in each of these studies was expected to be zero, it could be tested at the outset, and the result was $r = .12$, $p = .26$; the test of a quadratic relationship was also non-significant when *resactive*^{2} was added to *resactive* in a multiple regression (R^2 change = .007, p=.39).

3.1.1 Example 1 – *principles* regressed on *outclass*: test of theory

The first test addresses: "Does instructor availability outside of class have a positive relationship with students' understanding of principles and concepts?" The simple OLS solution was *predicted principles* = 44.230 + .404(*outclass*). For each additional increase of 1% in the percent of students who strongly agreed the instructor was available outside of class there was an increase of .404% in the percent of students who strongly agreed that they understood principles and concepts. This relationship was expected and is meaningful in terms of establishing availability as an instructor.

Now we ask: "Is *resactive* a suppressor variable for *outclass*"? *Resactive* was added, and the result was *predicted principles* = 43.582 + .398(*outclass*) + .386(*resactive*). The addition of *resactive* decreased the coefficient for *outclass* from .404 to .398, hence we reject our hypothesis that *resactive* is a suppressor for *outclass*.

3.1.2 Example 2 – *principles* regressed on *rgattend*: test of theory

The first test addresses: “Does regular class attendance have a positive relationship with students’ understanding of principles and concepts?” The OLS solution was *predicted principles* = 33.096 + .396(*rgattend*). For each additional increase of 1% in the percent of students who strongly agreed that regular class attendance was necessary there was an increase of .396% in the percent of students who strongly agreed that they understood principles and concepts. This relationship was expected and is meaningful in terms of the importance of attending class.

Now: “Is *resactive* a suppressor for *rgattend*”? *Resactive* was added, and the result was *predicted principles* = 23.813 + .460(*rgattend*) + 1.740(*resactive*). The addition of *resactive* increased the coefficient for *outclass* from .396 to .460, hence *resactive is* a suppressor variable. Since the correlation between *rgattend* and *resactive* was r = –.265 (p < .01) while *resactive* and *principles* were independent, the addition of *resactive* suppressed irrelevant variance in *rgattend* and magnified the importance of regular class attendance in understanding principles and concepts. In essence, the increased research activities did have a negative effect on the importance of attending class, but the students appear to have compensated for that by making the most of their time when the instructor was present.

Given that: a) the predictor *rgattend* and *resactive* are significantly correlated, b) *rgattend* is a significant predictor of *principles*, c) *resactive* and *principles* are not significantly correlated, and d) *resactive* is a suppressor for *rgattend*, we have an example of a suppressor variable that meets the definitions proposed by Mendershausen (1939), Horst (1941), and Pedhazur (1982, 1997).

The next question is whether the change in coefficients for *rgattend* from .396 to .460 is statistically significant. We apply the F-S test with $\tau = .396$, $\tau' = .460$, $\hat{\rho}_{X_1X_2} = -.265$, $\hat{\sigma}_{\tau} = .092$ and $\hat{\sigma}_{\tau'} = .093$. Using Equation 7, $t_{obs} = -2.585$. The t critical value for a one-tailed test with $\alpha = .05$ and $df \approx 100$ is –1.66. Since $-2.585\ t_{obs} < -1.66\ t_{cv}$, we reject the null hypothesis in favor of $|\tau| - |\tau'| < 0$ and conclude that the change in coefficients is statistically significant.

3.1.3 Example 3 – *principles* regressed on *time*: test of theory

The first test addresses: “Is the percent of time that students spend on the course more than others positively related to their understanding of principles and concepts”? The OLS solution was *predicted principles* = 53.506 + .180(*time*). For each additional increase of 1% in the percent of students who stated that they spent more or much more time on the course compared to others, there was an increase of .18% in the percent of students who strongly agreed that they understood principles and concepts. This relationship was expected and is meaningful in terms of the impact that students have on their own learning.

Now: “Is *resactive* a suppressor for *time*”? *Resactive* was added, and the result was *predicted principles* = 49.974 + .193(*time*) + 1.139(*resactive*). The addition of *resactive* increased the coefficient of the predictor *time* from .180 to .193, hence *resactive is* a suppressor. In contrast to Example 2, however, the correlation between the predictor *time* and suppressor *resactive* was not statistically significant (r = –.055, p = .567) yet the addition of *resactive* still clarified the strength of the relationship between *time* and *principles*. This example meets Pedhazur's (1982, 1997) definition of a suppressor but does not meet the definition of a suppressor according to Mendershausen (1939) and Horst (1941). As will be shown in the next section, this suppressor relationship would not have been discovered using Strategy #2 where potential suppressors are selected based on statistical criteria alone.

Similar to Example 2, we ask if the change in coefficients for *time* from .180 to .193 was statistically significant. The F-S test was applied with $\tau = .180$, $\tau' = .193$, $\hat{\rho}_{X_1X_2} = -.055$, $\hat{\sigma}_\tau = .060$, and $\hat{\sigma}_{\tau'} = .060$, and there was again a statistically significant change in the coefficients ($-3.938\ t_{obs} < -1.66\ t_{cv}$).

Note that $\tau - \tau'$ in Example 3 is less than $\tau - \tau'$ in Example 2 (.180 - .193 < .396 - .460) yet Example 3's t_{obs} is considerably larger: –3.938 versus –2.585. This apparent anomaly occurred because the standard errors in Example 3 (both are .060) are smaller than in Example 2 (where they are .092 and .093). The standard errors are smaller because the absolute value of the correlation in Example 3 between the suppressor and the predictor is smaller than Example 2 (|–.055| < |–.265|). This means there is less shared variance in the denominator of the standard error expressions that has to be adjusted for (see Equation 6). This relationship between collinearity in predictors and the magnitude of their standard errors illustrates one reason why textbooks recommend that predictors share as little variance as possible, i.e., the less shared variance, the smaller the standard errors, and the more powerful the t-test of the coefficients (Pedhazur 1997, p. 295).

3.2 Strategy #2 (Maximizing predictive validity variance)

The second strategy does not use theory to guide variable selection, but instead focuses on maximizing the variance explained by the predictor(s) of interest—we refer to this as the "predictive validity variance". This was first suggested by Horst (1941) who stated: "what we should do is systematically to investigate the variables whose *correlations with the criterion are negligible* in order to determine which of them have appreciable correlations with the prediction variables. Those which do should be included as suppression variables" (p. 435) (italics added). This recommendation defines an explicit *a priori* variable selection process for statistical, not necessarily substantive, purposes.

It is important to note that the predictive validity variance at issue here is not the overall $R^2_{y1\ldots k}$ for a k predictor problem. The variance that Horst refers to is just the variance accounted for by the predictor of interest since it is assumed that the variance accounted for by the suppressor is "negligible". Adapting Velicer's (1978) notation, the predictive validity variance accounted for in a k = 2 predictor model is the squared semi-partial (or part) correlation between the predictor and the criterion

Equation 8 $$r^2_{y(1.2)}$$

where *y*, 1, and 2 denote the criterion, predictor and suppressor, respectively. Hence, the strategy suggested by Horst involves the selection of any variable(s) that maximizes the relationship defined by Velicer (1978, Eq. 14)

Equation 9 $$r^2_{y(1.2)} > r^2_{y1}.$$

In essence this means that it is not additional variance in the criterion that is being accounted for through the introduction of additional criterion-correlated predictors, but rather, it is the relative proportion of variance the predictor of interest accounts for that is increased since "irrelevant" variance has been removed from the predictor by those variables that *act as* suppressors.

The following procedural steps were developed from Horst's definition: 1) select the criterion; 2) select the primary predictor of interest; 3) correlate the criterion with all variables other than the predictor and determine which of these variables are not significantly correlated with the criterion and consider them as potential suppressors; 4) correlate these potential suppressors with the predictor; and 5) use the variables that are significantly correlated with the predictor as suppressors.

To show the differences between Strategy #1 and Strategy #2, the same criterion (*principles*) and predictor (*time*) used in Example 3 from Strategy #1 are used for the next example. Instead of proposing hypotheses about suppressor variables, however, potential suppressor variables are selected using the statistical criteria laid out in steps 1 through 5 above. Since the same criterion and predictor are used, steps 1 and 2 are complete, and Example 4 begins with step 3.

3.2.1 Example 4 – *principles* regressed on *time*: maximize $\mathbf{r^2_{y(1.2)}}$

The correlation between *principles* and all other variables in the data set except *time* was determined and five variables were not significantly correlated with *principles* (step 3). These potential suppressors included: 1) *resactive*, the number of research activities; 2) *taught*, the number of times the course had been taught; 3) *tenure*, whether the professor was tenured or not; 4) *weekordr*, the order in the week the course was taught; and 5) *interact*, whether or not small-group interactions were incorporated in the course. The predictor *time* was then correlated with these five potential suppressors (step 4). One of them, *interact*, was significantly correlated with *time* ($r = –.594$, $p < .001$).

Interact was treated *as* a suppressor variable (step 5). *Time* was entered into the regression model followed by *interact*. The result was *predicted principles* = 31.391 + .291(*time*) + 12.315(*interact*), $r^2_{y(1.2)} = .132$. From Example 3 we found that *principles* regressed on *time* produced *predicted principles* = 53.506 + .180(*time*), $r^2_{y1} = .079$. From these results, we see that the addition of *interact* increased the coefficient of *time* from .180 to .291 and the explained validity variance from .079 to .132. Using Strategy #2 (maximizing $r^2_{y(1.2)}$), one variable out of the data set (*interact*) "*acts as*" a suppressor for *time* when predicting *principles*.

Unlike the three studies under Strategy #1, however, no attempt is necessary under this statistical strategy to provide a substantive explanation of why *interact* is suppressing irrelevant variance in *time*. Likewise, the F-S test is irrelevant because the goal of this strategy is simply to maximize the predictive validity variance by adding any potential suppressor correlated with the predictor but uncorrelated with the criterion. Note also that unlike Example 3 where *resactive* was hypothesized and found to be a suppressor variable for *time*, under the present statistical approach it was not identified and tested as a potential suppressor because *resactive* and *time* were not significantly correlated.

The one remaining observation to make about this approach concerns the point raised earlier by Wiggins (1973, p. 32), namely, is the incremental increase in predictive validity variance of .132 – .079 = .053 "practical"? In our experience with often messy social science survey data collected from relatively low reliability instruments, increments of 5% explained variance are meaningful and often difficult to achieve.

The zero-order correlations, covariances, means, and standard deviations for all variables in the preceding analyses are presented in Table 2. Readers familiar with the Sharpe and Roberts (1997) procedures may find it interesting to use the statistics in Table 2 to test other suppression situations that did not occur to us. Such activity will highlight the point of Strategy #2; namely, it is possible to conduct atheoretical analyses to maximize $r^2_{y(1.2)}$ and produce a suppression effect, but which offer no contribution to understanding the phenomenon under study without further investigation.

Table 2. Variable Means, Standard Deviations, Zero-Order Correlations, and Covariances

Covariances

Mean SD	1 n =105	2 n =105	3 n =105	4 n =110	5 n =110	6 n =110	7 n =110	8 n =110	9 n =110
1	64.50 18.53	205.64	132.19	151.22	5.51	16.38	.79	-.62	.12
2	.492	50.14 22.55	-33.91	26.97	7.61	-4.72	-.18	.50	-.19
3	.391	-.082	79.29 18.27	266.41	-12.24	27.21	.78	-1.74	-1.59
4	.282	.041	.504	58.94 29.89	-4.16	-22.05	-1.35	1.95	-7.77
5	.117	.133	-.265	-.055	2.32 2.52	.75	.19	-.27	.16
6	.147	-.035	.247	-.124	.050	8.76 5.95	.92	-2.19	.79
7	.125	-.024	.124	-.135	.222	.459	1.87 .34	-.10	.03
8	-.043	.029	-.123	.086	-.140	-.482	-.398	1.64 .76	-.07
9	.015	-.020	-.200	-.594	.142	.302	.223	-.215	1.25 .44

Correlations

Legend

1 *principles*; the percent of students in a class who strongly agreed that they understood principles and concepts

2 *outclass*; the percent of students in a class who strongly agreed that the instructor was available outside of class
3 *rgattend*; the percent of students in a class who strongly agreed that there was a necessity for regular class attendance
4 *time*; the percent of time spent on the class more than on other classes
5 *resactive*; the sum of workshops and seminars conducted, conference papers and invited presentations
6 *taught*; the number of times the course had been taught
7 *tenure*; whether the instructor was tenured or not
8 *weekordr*; the order in the week the course was taught
9 *interact*; whether or not small-group interactions were incorporated into the class

3.3 Strategy #3 (Post Hoc Determination)

Post hoc claims for suppression occur when researchers are not specifically looking for suppressor variables or effects, but conclude they may be present (e.g. Reeves and Pedulla 2011). This can occur in exploratory designs where variables are introduced into a regression model to simply see what criterion-predictor relationships exist—this is not uncommon in masters' theses and doctoral dissertations. It can also occur when the testing of an initial hypothesized model has generated unexpected results that the researcher feels compelled to explain somehow. In both situations, the existence of a suppressor was not initially considered, but through the course of subsequent analysis, a suppressor effect was found by chance.

The example below is an exploratory analysis in which familiarity with the data set suggests that multiple variables should provide a plausible test of sources of influence on the criterion – without any consideration of the extent to which any of them might *act as* suppressors. *Excell*, the percent of students who rated the instructor as excellent, is used as the criterion for this example instead of *principles*. *Excell* was selected because: 1) it is a meaningful indicator of how well students view an instructor; 2) it is useful to know what contributes to high instructor excellence ratings; and 3) *principles* has become too familiar to be used for this strategy.

3.3.1 Example 5 – *excell* regressed on *rgattend*, *principles*, *skills*, *outclass* and *time*: post hoc

As an exploratory analysis, it is interesting to look at variables that represent the students' experiences while in the course and whether those variables are useful predictors of the ratings they gave the instructor. For example, in these data students' experiences are captured through their ratings of: the necessity for regular class attendance (*rgattend*), their understanding of principles and concepts (*principles*), whether they believe they acquired academic skills (*skills*), the instructor's availability outside of class (*outclass*), and the percent of time spent on the class more than others (*time*).

These variables—*rgattend*, *principles*, *skills*, *outclass*, and *time*—were "force entered" as a block of predictors for the criterion *excell*. The result was *predicted excell* = –13.219 – .016(*rgattend*) + .799(*principles*) + .114(*skills*) + .162(*outclass)* – .124(*time*), $R^2 = .675$, $p < .001$. *Principles*, *outclass* and *time* are statistically significant ($p < .05$ for each), while *rgattend* and *skills* are not ($p > .2$ for each). For the practical purposes of an exploratory analysis, the analysis could stop, and various interpretations would be offered at this point.

However, a casual *post-hoc* inspection of the zero-order and partial correlations printed as part of the software output (Table 3) revealed an unexpected result. For *time*, the partial correlation with *excell* is considerably larger than its zero-order correlation ($|–.235| > .058$). This surprising finding is consistent with the definition of a suppression effect suggested by many researchers including Kendall and Stuart (1973, p. 330) and Pedhazur (1982).

Table 3. Zero-order and Partial Correlations from Output

Model	Zero-order Correlation	Partial Correlation
(Constant)		
rgattend	.196	-.018
principles	.783	.594
skills	.651	.113
outclass	.553	.217
time	.058	-.235

After next going through multiple iterations of changing the order and criteria in which these variables were entered into different models in order to determine which specific predictor(s) had affected the *time* estimates, we were forced to agree with Tzelgov and Henik (1991) who argue that in a multiple regression context "there is no simple way to identify suppressor variables" or the variable causing the suppression effect (p. 528). Similar to the situation with Example 4 (Strategy #2—maximizing predictive validity variance) no substantive explanation or interpretation of the suppressor effect is offered, and no F-S test is conducted. This specific example is the situation that prompted this paper because these results occurred when preparing the course lecture on suppressors, and no plausible explanation of the results was possible. The results were not expected, were not interpretable, and were ultimately attributed to chance, i.e., a complex statistical consequence of the collinear relationships existing within the predictor correlation matrix.

4. Conclusion

The primary purpose of this paper was to contribute to a greater practical understanding of suppressor variables for those who teach and use applied OLS multiple regression. Regression is a common statistical tool, and there are many routine but messy "dirty data" situations (e.g., low reliability measures, non-random missing data, weak interventions, poor sampling designs, and data entry errors) where one may not fully understand why results change when another predictor is simply added to one's model. For example, regardless of whether model building is for exploratory or theory testing purposes, the simple introduction of an additional predictor may diminish or enhance the strength of earlier regression coefficients. These changes are sometimes attributed to mediator or suppressor variables, but such attribution may suggest a causal mechanism that is unwarranted—particularly if theory is weak and data quality is suspect. Unfortunately, instruction in the use and identification of suppression is often confusing since there is no one universally employed definition of it—including when it has occurred, how to look for it, how to plan for its advantageous effects, or what statistical approach to use to test for and demonstrate its occurrence.

How then can we plan for suppressor variables, how do we detect suppression situations, how can we appropriately interpret their effects, and does it matter what kind of "effect" we call them? To answer these questions we find it useful to draw upon Kerlinger's (1964) distinction between explanation and prediction. His distinction provides a framework wherein two categories of statistical analyses for suppression may be identified: those based on a theoretical framework and those that are statistically based.

Within a theoretical framework we test hypotheses about whether or not a variable *is* a suppressor (Strategy #1). When suppressor analyses are not based on a theoretical framework, they are built to maximize predictive validity variance through the explicit addition of variables that may *act as* suppressors (Strategy #2). Finally, any regression model may be theoretically or statistically based, without any intention of discovering or testing suppression effects, yet simple or complex suppression effects may subsequently be observed by chance (Strategy #3).

Based on the three design and analysis strategies and five analytic examples illustrating their application, there are several points that may be taken away. First, each strategy has its own distinct purpose, set of conditions, and use. Hence, deliberation should be taken when considering possible suppressor variables, effects, and situations. Second, different strategies lead to different results most of the time (Table 1). This is because each approach has its own relatively unique procedure for defining when suppression has occurred. Finally, caution should be used when interpreting the results. Because of the differences in procedures, a claim that a variable *is* a suppressor and that is appropriate under one approach may not be appropriate for another—the warranty for such a claim depends on the strategy employed.

By presenting a history of suppressors and the confusion that surrounds them, as well as by comparing three strategies for working with suppressor variables and suppression effects, we hope that those who do regression modeling have extended their understanding of: a) the contrasting conceptual and definitional approaches, b) the underlying third-variable statistical relationships, c) how the results may be interpreted in different modeling situations, i.e. the claim a variable *is* a suppressor versus *acts as* a suppressor, and d) how the material might be taught in ways to highlight a-c. Ultimately, from our applied perspective, it is not *how* you do the testing for suppressors or suppression effects that matters but *why* you are doing the testing—in other words, the procedures follow the purpose.

References

Baron, R.M., & Kenny, D.A. (1986), “The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations,” *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bollen, K.A., & Stine, R. (1990), “Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability,” *Sociological Methodology*, 20, 115-140.

Burns, S., & Ludlow, L. H. (2005), “Understanding Student Evaluations of Teaching Quality: The Unique Contributions of Class Attendance,” *Journal of Personnel Evaluation in Education*. Available at http://dx.doi.org/10.1007/s11092-006-9002-7

Chapman, L., & Ludlow, L. H. (2010), “Can Downsizing College Class Sizes Augment Student Outcomes: An Investigation of the Effects of Class Size on Student Learning,” *Journal of*

General Education, 59(2), 105-123.

Cohen, J., & Cohen, P., (1975), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, New Jersey: Erlbaum.

Conger, A. J. (1974), "A Revised Definition for Suppressor Variables: A Guide to Their Identification and Interpretation," *Educational and Psychological Measurement*, 34(1), 35-46.

Cronbach, L.J. (1950), "Further Evidence on Response Sets and Test Design," *Educational and Psychological Measurement*, 10, 3-31.

Currie, I., & Korabinski, A. (1984), "Some Comments on Bivariate Regression," *The Statistician*, 33, 283-293.

Darlington, R. B. (1968), "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin*, 69(3), 161-182.

Dearing, E., & Hamilton, L.C. (2006), V., "Contemporary Advances and Classic Advice for Analyzing Mediating and Moderating Variables," *Monographs of the Society for Research in Child Development*, 71: 88–104. doi: 10.1111/j.1540-5834.2006.00406.x

Dicken, C. (1963), "Good Impression, Social Desirability and Acquiescence as Suppressor Variables," *Educational and Psychological Measurement*, 23, 699-720.

Efron, B., & Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall/CRC.

Field, A. (2013), *Discovering Statistics Using IBM SPSS Statistics*, 4th Edition, Los Angeles: SAGE.

Fox, M. F. (1992), "Research, Teaching, and Publication Productivity: Mutuality versus Competition in Academia," *Sociology of Education*, 65, 293-305.

Freedman, L. S., & Schatzkin, A. (1992), "Sample Size for Studying Intermediate Endpoints within Intervention Trials of Observational Studies," *American Journal of Epidemiology*, 136, 1148-1159.

Friedman, L., & Wall, M. (2005), "Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression," *The American Statistician*, 59(2), 127-136.

Gulliksen, H. (1950), *Theory of Mental Tests*, New York, NY: Wiley.

Hamilton, D. (1987), "Sometimes $R^2 > r_{yx_1}^2 + r_{yx_2}^2$," *The American Statistician*, 41, 129-132.

Hamilton, D. (1988), (Reply to Freund and Mitra), *The American Statistician*, 42, 90-91.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003), *Applied Statistics for the Behavioral Sciences*, 5th edition, Boston, MA: Houghton Mifflin Company.

Holling, H. (1983), "Suppressor Structures in the General Linear Model," *Educational and Psychological Measurement*, 43(1), 1-9.

Horst, P. (1941), "The Prediction of Personal Adjustment," *Social Science Research Council Bulletin*, 48.

Iacobucci, D. (2008), *Mediation Analysis*, Los Angeles: SAGE.

Judd, C.M., & Kenny, D.A. (1981), "Process Analysis: Estimating Mediation in Treatment Evaluations," *Evaluation Review*, 5, 602-619.

Kendall, M. G., & Stuart, A. (1973), *The Advanced Theory of Statistics* (Vol. 2, 3rd ed.), New York: Hafner Publishing.

Kenny, D.A., Kashy, D.A., & Bolger, N. (1998), "Data Analysis in Social Psychology." In D. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology* (4th ed., Vol. 1, pp. 233-265). New York: McGraw-Hill.

Kerlinger, F. N. (1964), *Foundations of Behavioral Research*, New York, NY: Holt, Rinehart and Winston, Inc.

Kerlinger, F. N., & Pedhazur, E. J. (1973), *Multiple Regression in Behavioral Research*, New York, NY: Holt, Rinehart and Winston, Inc.

Logan, J. A. R., Schatschneider, C., & Wagner, R. K. (2009), "Rapid Serial Naming and Reading Ability: The Role of Lexical Access," *Reading and Writing*, 24, 1-25.

Lord, F. M., & Novick, M. R. (1968), *Statistical Theories of Mental Test Scores*, Reading, MA: Addison-Wesley Publishing Company, Inc.

Lubin, A. (1957), "Some Formulae for Use with Suppressor Variables," *Educational and Psychological Measurement*, 17(2), 286-296.

Ludlow, L. H. (1996), "Instructor Evaluation Ratings: A Longitudinal Analysis," *Journal of Personnel Evaluation in Education*, 10, 83-92.

Ludlow, L. H. (2002), "Rethinking Practice: Using Faculty Evaluations to Teach Statistics," *Journal of Statistics Education*, 10(3). Available at www.amstat.org/publications/jse/v10n3/ludlow.html

Ludlow, L. H. (2005), "A Longitudinal Approach to Understanding Course Evaluations," *Practical Assessment Research and Evaluation*, 10(1), 1-13. Available at http://pareonline.net/pdf/v10n1.pdf

Ludlow, L. H., & Alvarez-Salvat, R. (2001), "Spillover in the Academy: Marriage Stability and Faculty Evaluations," *Journal of Personnel Evaluation in Education*, 15(2), 111-119.

Ludlow, L.H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J.B. (2014), "Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering," *Measurement and Evaluation in Counseling and Development*. 47(2), 127-149.

Lynn, H. S. (2003), "Suppression and Confounding in Action," *The American Statistician*, 57(1), 58-61.

Maassen, G. H., & Bakker, A. B. (2001), "Suppressor Variables in Path Models: Definitions and Interpretations," *Sociological Methods & Research*, 30(2), 241-270.

MacKinnon, D. P. (2008), *Introduction to Statistical Mediation Analysis*, New York: LEA.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000), "Equivalence of the Mediation, Confounding and Suppression Effect," *Prevention Science*, 1(4), 173-181.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002), "A Comparison of Methods to Test Mediation and Other Intervening Variable Effects," *Psychological Methods*, 7(1), 83-104.

MacKinnon, D.P., Lockwood, C.M., & Williams, J. (2004), "Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods," *Multivariate Behavioral Research*. 2004 January 1; 39(1): 99. doi: 10.1207/s15327906mbr3901_4

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995), "A Simulation Study of Mediated Effect Measures," *Multivariate Behavioral Research*, 30(1), 41-62.

McClendon, W. J. (1994), *Multiple Regression and Causal Analysis*, Itasca, Il: F.E. Peacock Publishers, Inc.

McFatter, R. M. (1979), "The Use of Structural Equation Models in Interpreting Regression Equations Including Suppressor and Enhancer Variables," *Applied Psychological Measurement*, 3, 123-135.

McNemar, Q. (1945), "The Mode of Operation of Suppressant Variables." *American Journal of Psychology*, 58, 554-555.

McNemar, Q. (1949), *Psychological Statistics*, New York, NY: Wiley.

Meehl, P.E. (1945). "A Simple Algebraic Development of Horst's Suppressor Variables," *American Journal of Psychology*, 58, 550-554.

Mendershausen, H. (1939), "Clearing Variates in Confluence Analysis," *Journal of the American Statistical Association*, 34(205), 93-105.

Pandey, S., & Elliott, W. (2010), "Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models," *Journal of the Society for Social Work and Research*, 1(1), 28-40.

Pedhazur, E. J. (1982), *Multiple Regression in Behavioral Research: Explanation and Prediction*, 2nd edition, Fort Worth, TX: Harcourt Brace College Publishers.

Pedhazur, E. J. (1992), "In Memoriam – Fred N. Kerlinger (1910-1991)," *Educational Researcher*, 21(4), 45.

Pedhazur, E. J. (1997), *Multiple Regression in Behavioral Research: Explanation and Prediction*, 3rd Edition, Fort Worth, TX: Harcourt Brace College Publishers.

Preacher, K. J., & Hayes, A. F. (2008), "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models," *Behavior Research Methods*, 40(3), 879-891.

Ramsden, P., & Moses, I. (1992), "Association between Research and Teaching in Australian Higher Education," *Higher Education*, 23, 273-295.

Reeves, T. D., & Pedulla, J. J. (2011), "Predictors of Teacher Satisfaction with Online Professional Development: Evidence from the USA's e-Learning for Educators Initiative," *Professional Development in Education*, 37(4), 1-21.

Schey, H. M. (1993), "The Relationship between the Magnitudes of SSR(x_2) and SSR($x_2|x_1$): A Geometric Description," *The American Statistician*, 47, 26-30.

Sharpe, N. R., & Roberts, R. A. (1997), "The Relationship Among Sums of Squares, Correlation Coefficients, and Suppression," *The American Statistician*, 51(1), 46-48.

Shieh, G. (2001), "The Inequality Between the Coefficient of Determination and the Sum of Squared Simple Correlation Coefficients," *The American Statistician*, 55(2), 121-124.

Shieh, G. (2006), "Suppression Situations in Multiple Linear Regression," *Educational and Psychological Measurement*, 66(3), 435-447.

Shrout, P.E., Bolger, N. (2002), "Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendation," *Psychological Methods*, 7(4), 422-445.

Smith, R. L. Ager, J. W., Jr., & Williams, D. L. (1992), "Suppressor Variables in Multiple Regression/Correlation," *Educational and Psychological Measurement*, 52, 17-29.

Sobel, M. E. (1982), "Asymptotic Confidence Intervals for Indirect Effects in Structural

Equation Models," In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington, DC: American Sociological Association.

Sobel, M.E. (1986), "Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models." In N. Tuma (Ed.), *Sociological Methodology 1986* (pp. 159-186). Washington, DC: American Sociological Association.

Spearman, C. (1904), "The Proof and Measurement of Association between Two Things," *American Journal of Psychology*, 15, 72-101.

Stewart, P. W., Reihman, J., Lonky, E., & Pagano, J. (2012), "Issues in the Interpretation of Associations of PCBs and IQ," *Neurotoxicology and Teratology*, 34, 96-107.

Tu, Y. K., Gunnell, D., & Gilthorpe, M. S. (2008), "Simpson's Paradox, Lord's Paradox, and Suppression Effects are the Same Phenomenon – the Reversal Paradox," *Emerging Themes in Epidemiology*, 5(2).

Tzelgov, J., & Henik, A. (1985), "A Definition of Suppression Situations for the General Linear Model: A Regression Weights Approach," *Educational and Psychological Measurement*, 45(2), 281-284.

Tzelgov, J., & Henik, A. (1991), "Suppression Situations in Psychological Research: Definitions, Implications, and Applications," *Psychological Bulletin*, 109(3), 524-536.

Tzelgov, J., & Stern, I. (1978), "Relationships between Variables in Three Variable Linear Regression and the Concept of Suppressor," *Educational and Psychological Measurement*, 38(2), 325-335.

Velicer, W. F. (1978), "Suppressor Variables and the Semipartial Correlation Coefficient," *Educational and Psychological Measurement*, 38(4), 953-958.

Voyer, D. (1996), "The Relation between Mathematical Achievement and Gender Differences in Spatial Abilities: A Suppression Effect," *Journal of Educational Psychology*, 88(3), 563-571.

Voyer, D., & Sullivan, A.M. (2003), "The Relation between Spatial and Mathematical Abilities: Potential Factors Underlying Suppression," *International Journal of Psychology*, 38(1), 11-23.

Walker, D. A. (2003), "Suppressor Variable(s) Importance within a Regression Model: An Example of Salary Compression from Career Services," *Journal of College Student Development*, 44(1), 127-133.

Wherry, R.J. (1946), "Test Selection and Suppressor Variables." *Psychometrika*, 11, 239-247.

Wiggins, J.S. (1973), *Personality and Prediction: Principles of Personality Assessment.* Reading, MA: Addison-Wesley, 1973.

Larry Ludlow
Boston College
Lynch School of Education
Educational Research, Measurement, and Evaluation Department
140 Commonwealth Avenue
336C Campion Hall
Chestnut Hill, MA 02467
USA
Phone: 617-552-4221
ludlow@bc.edu

Kelsey Klein
Boston College
Lynch School of Education
Educational Research, Measurement, and Evaluation Department
140 Commonwealth Avenue
336C Campion Hall
Chestnut Hill, MA 02467
USA
kleinkg@bc.edu
