*Communication Methods and Measures*, 6:1–11, 2012

ISSN: 1931-2458 print/1931-2466 online
DOI: 10.1080/19312458.2012.651415

# EDITORIAL

# Cautions Regarding the Interpretation of Regression Coefficients and Hypothesis Tests in Linear Models with Interactions

Andrew F. Hayes, Carroll J. Glynn, and Michael E. Huge
*School of Communication, The Ohio State University*

Manuscripts submitted to *Communication Methods and Measures* and most empirical journals in the field of communication frequently rely on some kind of linear model in the data analysis. Most of the more popular statistical methods can be framed in terms of a linear model, including the independent and dependent groups *t* tests, analysis of variance and covariance, multiple regression, multilevel modeling, and structural equation modeling. Of these methods, multiple regression and its special cases, such as analysis of variance and covariance (Cohen, 1968), are by far the most frequently used. It is safe to say that in order to understand and publish in the empirical communication literature, familiarity with the basic principles of linear modeling is a necessity.

While the use and interpretation of a linear model is standard curriculum in research-oriented graduate programs in communication, a solid understanding of the versatility and subtle complexities of linear modeling requires more than just a week or two devoted to the topic in an introductory data analysis course. The ability for variables in a linear model to influence each other's effects is a case in point. When two variables are allowed to influence each other's effects, meaning that two variables *interact*, the interpretations of some of the coefficients in a linear model are quite different compared to when interaction effects are not included. The additional complexities and potential for misinterpretation are so great that whole chapters of statistical methods texts (e.g., Cohen, Cohen, West, & Aiken, 2003, chapters 7 and 9; Darlington, 1990, chapter 13; Hayes, 2005, chapter 16) and even entire books are devoted to the topic of interactions in multiple regression (e.g., Aguinis, 2003; Aiken & West, 1991; Jaccard & Turrisi, 2003; Kam & Franzese, 2007).

Correspondence should be addressed to Andrew F. Hayes, School of Communication, The Ohio State University, 3016 Derby Hall, 154 N.Oval Mall, Columbus OH 43210. E-mail: hayes.338@osu.edu

Unfortunately, our casual examination of the field's journals (as well as this editor's experience with some manuscripts submitted to *Communication Methods and Measures*) tells us that some of the important details about the interpretation of regression models with interactions are not well understood by communication researchers, reviewers, and journal editors. As a result, there are published instances where investigators misinterpreted the coefficients in a linear model with interactions by overgeneralizing their understanding of the concept of a "main effect" from analysis of variance and covariance to any linear model. In this editorial primer, we analytically describe and illustrate by way of an example that when two variables are allowed to interact in a multiple regression model, the coefficients in the model for those variables are *conditional effects* and not main effects, as they often are misinterpreted. We show that the size of the coefficients for variables allowed to interact in a linear model (and the results of hypothesis tests thereof) is influenced by decisions about the scaling of those variables that frequently are made arbitrarily. Although we illustrate in the context of ordinary least squares regression, these lessons apply to any linear model, including logistic regression, multilevel models, Poisson regression and other linear models of counts, ordinal regression, and so forth.

The points made here are not original by any means. There are numerous articles in the literature that make these same points (Bauer & Curran, 2005; Friedrich, 1982; Hayes & Matthes, 2009; Irwin & McClelland, 2001; Whisman & McClelland, 2005), and the lessons described here are also described clearly in the books frequently cited by communication scholars reporting a regression analysis (e.g., Aiken & West, 1991; Cohen et al., 2003; Jaccard & Turrisi, 2003) and therefore presumably read by those same scholars. Yet neither are these points hidden away in footnotes nor tucked in obscure technical appendices in those books. We believe the principles we discuss here at not sufficiently understood by many and so are worth repeating in the field's methodology journal. Our primary goal here is to further their dissemination in the hopes of improving the accuracy of the inferences and claims researchers make from their data, both in papers submitted to *Communication Methods and Measures* and elsewhere.

## SOME GENERAL INTERPRETATIONAL PRINCIPLES

A linear model is an additive combination of variables, each given some weight in an equation of the outcome variable, with the weights for each variable estimated from the data such that the resulting model "best fits" the data, with the ordinary least squares criterion being the most common approach to maximizing fit when $Y$ is treated as a continuum. Such a linear model takes the form

$$\hat{Y} = i + b_1X + b_2M + \sum_{j=1}^{k} g_jC_j \tag{1}$$

where $\hat{Y}$ is the expected value on the outcome variable, $X$ and $M$ are predictor variables we focus on in the discussion that follows, and $C_j$, $j = 1$ to $k$, represent $k$ covariates each weighted by a partial regression coefficient $g_j$.[1] In this model, $b_1$ quantifies the expected difference in

[1]The interpretational principles we describe throughout this manuscript apply to models without covariates as well. We include them here for generality.

$Y$ associated with a one unit difference in $X$ while holding $M$ and the $k$ covariates constant. Rephrased, $b_1$ measures how $Y$ is expected to change as $X$ changes by one unit, all else being held constant. The hypothesis test associated with $b_1$ (i.e., the $t$ and $p$-value) is used to test the null hypothesis that there is no linear partial relationship between $Y$ and $X$.

In a model of this form, whether one chooses $M = 2$, $C_1 = 5$, and $C_2 = 6$, for instance, or $M = 5$, $C_1 = -2$, and $C_2 = 7$, the expected difference in $Y$ between two cases who differ by one unit on $X$ remains $b_1$. Thus, in causal language, we can say that $X$'s effect on $Y$ is *conditionally invariant* because it does not depend on any other variable in the model. Similarly, $b_2$ quantifies the expected difference in $Y$ associated with a one unit difference in $M$, holding $X$ and the covariates constant. Rephrased, $M$'s effect on $Y$ is conditionally invariant, such that changing $M$ by one unit corresponds with an expected change in $Y$ of $b_2$ units, regardless of the values of $X$ and the covariates. These interpretations apply to standardized regression coefficients as well (i.e., when measurements on the variables are expressed as standard deviations from the sample mean), in which case "one unit" refers to one standard deviation of the standardized $X$ and $M$, and $b_1$ and $b_2$ represent the expected difference in $Y$ in standard deviations.

Conditional invariance is often assumed when communication researchers estimate regression models to test hypotheses. However, researchers often hypothesize that a variable's effect is not conditionally invariant, meaning, for example, that $X$'s effect depends on $M$. If $X$'s effect depends on $M$, then we say that $X$ and $M$ *interact*, or that $M$ *moderates* the effect of $X$.

Although there are many ways such interactions can be set up in a linear model, linear interaction is the most widely used. In such a model, $X$'s effect is allowed to vary as a function of $M$ by constructing a new variable defined as the product of $X$ and $M$ and including it in the model along with $X$ and $M$, as such:

$$\hat{Y} = i + b_1X + b_2M + b_3XM + \sum\nolimits_{j=1}^{k} g_jC_j \tag{2}$$

This model relaxes the conditional invariance constraint by allowing the regression weight for $X$ to vary linearly as a function of $M$. That the effect of $X$ is a function of $M$ is most easily illustrated by a simple algebraic manipulation of Equation 2:

$$\hat{Y} = i + (b_1 + b_3M)X + b_2M + \sum\nolimits_{j=1}^{k} g_jC_j \tag{3}$$

Notice here that $X$'s effect is $b_1 + b_3\,M$ and so depends on $M$. The *symmetry property* of interactions (Darlington, 1990) tells us that $X$ can also be construed as a moderator of $M$, as an equivalent version of Equation 2 after algebraic manipulation illustrates:

$$\hat{Y} = i + b_1X + (b_2 + b_3X)M + \sum\nolimits_{j=1}^{k} g_jC_j \tag{4}$$

Observe in Equation 4 that $M$'s effect on $Y$, $b_2 + b_3\,X$, is a function of $X$.

In Equations 2 and 3, $b_3$ quantifies how much the effect of a one-unit change in $X$ on $Y$ itself changes when $M$ changes by one unit. Conversely, $b_3$ quantifies how much the effect of a one-unit change in $M$ on $Y$ changes when $X$ changes by one unit (as Equation 4 illustrates). If $b_3$ is different from zero to a statistically significant degree, this implies that $X$'s effect on $Y$ is not conditionally invariant across values of $M$, or, conversely, that $M$'s effect on $Y$ is not conditionally invariant across values of $X$. That is, $X$ and $M$ interact.

## CONDITIONAL EFFECTS VERSUS MAIN EFFECTS

The most common misunderstanding in the interpretation of a model of the form of Equation 2 is the meaning of $b_1$ and $b_2$ and their corresponding tests of significance. Likely because of widespread familiarity with interactions in the context of analysis of variance and covariance, these coefficients are often interpreted analogous to "main effects" in analysis of variance, meaning an average effect of $X$ (for $b_1$, or $M$ for $b_2$) when collapsing across the other variable involved in the interaction.[2] But typically they do not have such an interpretation, and talking about them as such is inappropriate. Most generally, $b_1$ represents the expected difference in $Y$ associated with a one-unit difference in *$X$ when $M$ equals zero.* As such, $b_1$ represents a *conditional effect* of $X$, or, in the language of analysis of variance, a *simple effect*. Similarly, $b_2$ is the expected difference in $Y$ associated with a one-unit difference in *$M$ when $X$ equals zero*. It too is a conditional or simple effect rather than a main effect. The tests of significance for $b_1$ and $b_2$ test whether there is evidence that the conditional effect is different from zero.

The interpretation of $b_1$ as $X$'s effect conditional on $M = 0$ can be seen most clearly in Equation 2 by noticing that if $M$ is set to 0, Equation 2 reduces to $\hat{Y} = a + b_1X + \sum_{j=1}^{k} \gamma_j C_j$. In this case, $Y$ is estimated to change by $b_1$ units as $X$ changes by one unit, as nothing else in the model links $X$ to $Y$ except $b_1$. But if $M$ is some value other than 0, the expected change in $Y$ associated with a one-unit change in $X$ is not $b_1$ but, instead, $b_1 + b_3 M$. So $X$'s effect on $Y$ is $b_1$ only when $M = 0$ (or, alternatively, when $b_3 = 0$). A similar reasoning leads to the interpretation that $M$'s effect on $Y$ is $b_2 + b_3 X$. So $b_2$ represents the effect of a one-unit change in $M$ on $Y$ only when $X = 0$ or $b_3 = 0$.

Figure 1 graphically represents $b_1$ and $b_2$ in a regression model of the form in Equation 2. In this figure, it is clear why Darlington (1990) calls $b_1$ and $b_2$ *local terms* of the model, as they represent the relationship between $X$ and $Y$, or $M$ and $Y$, in specific regions of the regression surface. By contrast, $b_3$ is a *global term* of the model, in that it is a characteristic of the entire regression solution—specifically, the "warp" in the regression surface.

## IMPLICATIONS OF THE MISUNDERSTANDING

The interpretation of $b_1$ and $b_2$ as conditional effects in a model of the form represented in Equation 2 must be understood when discussing the results from a regression model. When investigators make decisions about the scaling or coding of predictors hypothesized to moderate the effects of other variables—decisions that frequently are arbitrary—those decisions can influence the size and statistical significance of some of the coefficients in the regression model. A failure

[2]In ANOVA, a main effect is the unweighted average simple effect of one variable, averaging across levels of the second. This concept does not generalize to regression models with interactions involving quantitative predictor variables. See Darlington (1990, pp. 440–441) for a discussion. Hayes (2005, p. 440–441) shows an example where the conditional effects in a regression model are in fact equivalent to main effects in a $2 \times 2$ ANOVA.

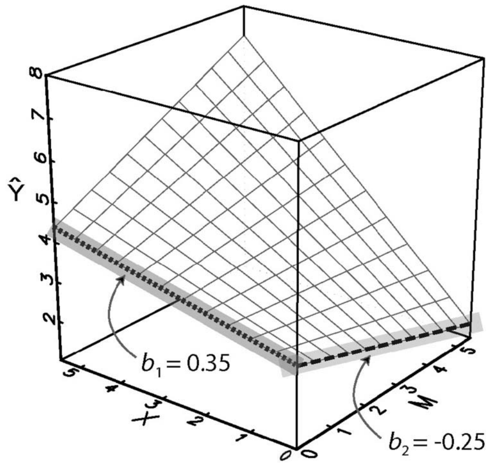


FIGURE 1 A graphical depiction of $\hat{Y} = i + b_1X + b_2M + b_3XM$ when $i = 3$, $b_1 = 0.35$, $b_2 = -0.25$, $b_3 = 0.15$.

to understand the effects of those decisions can yield misinterpretations of those coefficients, and, therefore, whether a substantive claim is actually supported by the statistical evidence.

To illustrate, we use data from a study published in the *Journal of Broadcasting and Electronic Media* (Glynn, Huge, Reineke, Hardy, & Shanahan, 2007) where the very interpretational oversights described here were made yet were not caught by the reviewers or the editor. These data come from several hundred phone interviews of a national probability sample of U.S. adults. Respondents were asked a number of questions about the extent of their support for government involvement in family issues, such as expanding Medicare coverage, government-funded medical leave and day care, and increased education spending. Responses were made on a 1 (strongly oppose) to 5 (strongly support) scale, yielding a scale of "support for social spending" that could range between 1 and 5. Participants were also asked questions to measure "perceived reality" of daytime talk shows (such as Oprah Winfrey, Ricki Lake, Rosie O'Donnell, and so forth) with respect to discussion of controversial social issues. An example question is "How accurately do you think talk shows represent issues that are important?" Responses were aggregated to form a scale between 1 and 5, with higher scores reflecting greater perceived reality of daytime talk shows. Participants were also asked their political ideology on a 1 (extremely conservative) to 7 (extremely liberal) scale and various demographics and other measures not pertinent to our discussion here.

Glynn et al. (2007) hypothesized that the greater the perceived reality of daytime talk shows, the more the respondent would express support for government social spending ($H_2$ in their paper). They also proposed that the relationship between political ideology and support for government social spending would depend on perceived reality of daytime talk shows ($H_4$ in their paper). We will not contradict any of their claims here, but merely point out that the evidence used to support some of the claims made are not pertinent due to their misinterpretation of $b_1$ and $b_2$ in a model of the form in question 2, as they estimated.

Using the Glynn et al. (2007) data, the first column in Table 1 presents the coefficients from an ordinary least squares regression model of the form

TABLE 1
OLS Regression Coefficients, Inferential Statistics, and Measures of Effect Size for Perceived Reality and Political Ideology as a Function of Scaling Choices in a Moderated Multiple Regression

| | *Model 1* | *Model 2* | *Model 3* | *Model 4* | *Model 5* | *Model 6* |
|---|---|---|---|---|---|---|
| | *No interaction in model* | *Original data (reported in Glynn et al.)* | *Ideology reverse scored* | *Ideology rescaled-3 to 3* | *Ideology and Reality mean centered* | *Ideology and Reality standardized* |
| Reality | | | | | | |
| $b_1$ | 0.090 | 0.503 | −0.332 | 0.086 | 0.099 | 0.069 |
| se($b_1$) | 0.047 | 0.130 | 0.133 | 0.046 | 0.047 | 0.032 |
| $t$ | 1.917 | 3.851 | −2.494 | 1.846 | 2.125 | 2.125 |
| $p$ | 0.056 | <.001 | 0.013 | 0.066 | 0.034 | 0.034 |
| $\beta$ | 0.082 | 0.460 | −0.303 | 0.078 | 0.090 | 0.090 |
| $sr^2$ | 0.006 | 0.023 | 0.009 | 0.005 | 0.007 | 0.007 |
| $pr^2$ | 0.007 | 0.029 | 0.013 | 0.007 | 0.009 | 0.009 |
| Ideology | | | | | | |
| $b_2$ | 0.160 | 0.421 | −0.421 | 0.421 | 0.163 | 0.229 |
| se($b_2$) | 0.022 | 0.080 | 0.080 | 0.080 | 0.022 | 0.030 |
| $t$ | 7.338 | 5.255 | −5.255 | 5.255 | 7.555 | 7.555 |
| $p$ | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 |
| $\beta$ | 0.296 | 0.778 | −0.778 | 0.778 | 0.302 | 0.302 |
| $sr^2$ | 0.084 | 0.042 | 0.042 | 0.042 | 0.087 | 0.087 |
| $pr^2$ | 0.099 | 0.053 | 0.053 | 0.053 | 0.104 | 0.104 |
| Ideology × Reality | | | | | | |
| $b_3$ | | −0.104 | 0.104 | −0.104 | −0.104 | −0.102 |
| se($b_3$) | | 0.031 | 0.031 | 0.031 | 0.031 | 0.030 |
| $t$ | | −3.384 | 3.384 | −3.384 | −3.384 | −3.384 |
| $p$ | | <001 | <.001 | <.001 | <.001 | <.001 |
| $\beta$ | | −0.647 | 0.611 | −0.500 | −0.132 | −0.132 |
| $sr^2$ | | 0.017 | 0.017 | 0.017 | 0.017 | 0.017 |
| $pr^2$ | | 0.023 | 0.023 | 0.023 | 0.023 | 0.023 |
| $R^2$ | 0.235 | 0.252 | 0.252 | 0.252 | 0.252 | 0.252 |

$b$ = unstandardized regression coefficient; $\beta$ = standardized regression coefficient; $sr^2$ = squared semipartial correlation; $pr^2$ = squared partial correlation. $sr^2$ is equivalent to "change in $R^2$" when the predictor is added to a model containing the other predictors. $pr^2$ is also known as partial $\eta^2$.

$$\hat{Y} = i + b_1 REALITY + b_2 IDEO + \sum_{j=1}^{k} g_j C_j$$

where $\hat{Y}$ is estimated support for social spending, *REALITY* and *IDEO* are perceived reality and political ideology, respectively, and $C_j$ refers to each of $k$ control variables, including exposure to daytime talk shows, income, gender, education, age, and race (white versus not white). Because they are not pertinent to our discussion, we exclude the coefficients for the control variables (the $k$ values of $g_j$ in the model above) from Table 1. As can be seen in the column labeled "Model 1", controlling for political ideology and all the covariates, there is a positive partial association between perceived realism of daytime talk shows and support for social spending. That is, two people of the same ideology, education, race, income, age, and sex, and who are equally exposed to daytime talk shows, but who differ by one unit in their perceptions of the reality of those shows, are expected to differ by 0.090 units in their support for social spending. The positive sign of the coefficient reflects the tendency for people who perceive greater realism to be more supportive of social spending. However, this coefficient just misses statistical significance ($p = 0.056$). This is largely consistent with hypothesis 2 in Glynn et al. (2007), depending on one's comfort with the concept of "marginally significant."

However, this is not the model that Glynn et al. (2007) reported. Rather, they estimated and report the coefficients from a model like Equation 2:

$$\hat{Y} = i + b_1 REALITY + b_2 IDEO + b_3(REALITY \times IDEO) + \sum\nolimits_{j=1}^{k} g_j C_j$$

The corresponding coefficients can be found under "Model 2" in Table 1, along with various other measures of effect such as the standardized regression coefficient, the squared semipartial correlation (conceptually similar to $\eta^2$ or the "change in $R^2$" using hierarchical entry—the proportion of the total variance in support for social spending that is explained uniquely by that predictor with all other variables in the model) and the squared partial correlation (conceptually similar to partial $\eta^2$—the proportion of the variance in support for social services left unexplained by other variables in the model that is uniquely explained by that predictor). Based on this model, Glynn et al. (2007) claim support for hypothesis 2 on the grounds that the coefficient for perceived reality is positive and statistically significant. However, as discussed above, $b_1 = 0.503$ estimates the relationship between perceived reality and support for social spending *conditional on political ideology equaling zero* (see Figure 2 for a graphical depiction) and the corresponding $t$ and $p$-value is used to test the null hypothesis that this conditional effect equals zero. But recall from the measurement of political ideology described earlier that zero is outside of the bounds of the measurement scale, as ideology is scaled to be between 1 and 7. Therefore, $b_1 = 0.503$ and its $p$-value, at best, represents an extrapolation beyond the data and requires a leap of faith in what the model would have looked like had respondents been given the opportunity to report that they were more than "extremely" conservative (which is the lowest point on the scale, arbitrarily set to 1 in terms of the measurement scheme).

Just as $b_1$ represents the conditional effect of perceived reality given political ideology equals zero, $b_2$ represents the conditional effect of political ideology given perceived reality equals zero. Glynn et al. (2007) report, based on $b_2 = 0.421$, $p < .001$ in Table 1, that this supports their expectation (though not explicitly hypothesized) that liberals would be more supportive of social spending than conservatives. Again, this represents the relationship between ideology and support *when perceived reality is zero*. As with political ideology, a perceived reality value of zero is outside the bounds of the measurement scale (as it was bound between 1 and 5), and thus this coefficient and its $p$-value cannot be used to substantiate this claim. To their credit, Glynn et al. do not emphasize this finding in light of $H_4$, which proposed that the relationship between ideology and support for social spending would depend on perceived reality of daytime talk shows, an interaction which was statistically significant.

Returning to hypothesis 2, if $b_1 = 0.503$ was in fact a global property of the model rather than a local property, then it should not be affected by a simple additive transformations of political ideology, such as subtracting a constant from each respondent's score. For example, imagine that

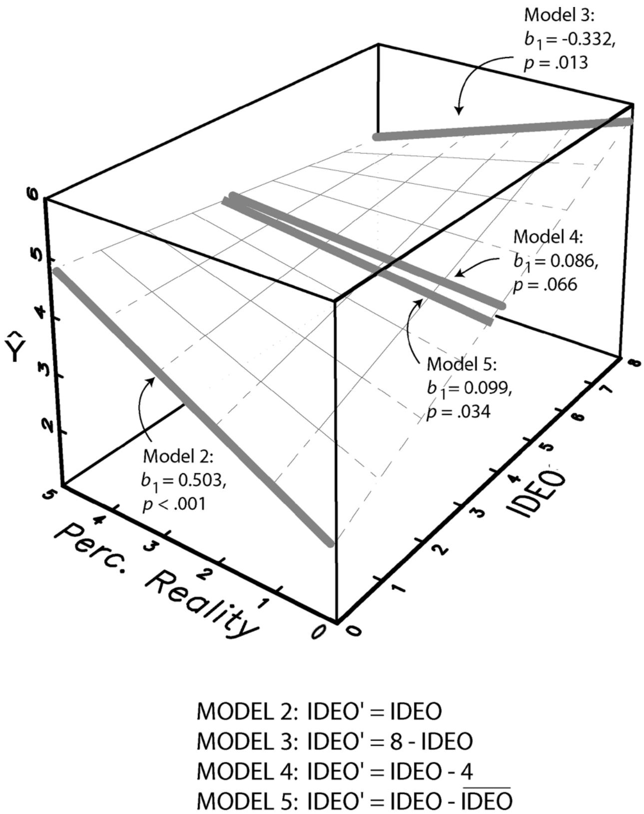


FIGURE 2 A graphical depiction of $b_1$ in Models 2, 3, 4, and 5 from Table 1.

rather than coding political ideology such that 1 = extremely conservative and 7 = extremely liberal, this scale was coded such that 7 = extremely conservative and 1 = extremely liberal. This is equivalent to subtracting the original scores from 8 (i.e., IDEO′ = 8 – IDEO) and then reeestimating the model substituting IDEO′ for IDEO throughout. The coefficients from the model after this simple transformation can be found in Table 1, Model 3. Notice that this arbitrary decision about the scaling of political ideology has dramatically changed the coefficient for perceived reality from significantly positive to significantly *negative*, $b_1 = -0.332$, $p = 0.013$, which would lead to the misinterpretation that perceptions of greater realism are associated with reduced support for social spending (see Figure 2 for a graphical depiction). However, this is perfectly consistent with the logic of the conditional interpretation reported earlier. In this model, $b_1$ represents a projection beyond the data and a leap of faith in what the relationship between perceived support and perceived realism would be for those participants who, had they been giving the opportunity, reported that they were even more liberal than "extremely" liberal (i.e., a hypothetical zero on a new 0-7 scale, where 1 = "extremely liberal").

An equally arbitrary scaling decision for political ideology could have been implemented, such as scaling ideology between -3 and 3 (where -3 = extremely conservative and 3 = extremely liberal), which is equivalent mathematically to a transformation of the form IDEO$'$ = IDEO – 4. The “Model 4” column of Table 1 shows the coefficients after this transformation. Notice that $b_1$ is different still, not quite statistically significant, but positive as predicted, $b_1 = 0.086$, $p = 0.066$. Now $b_1$ estimates the conditional effect of perceived realism for those responding right in the middle of the scale (as 0 would be the value recorded for the middle response option on a seven point -3 to 3 scale). See Figure 2 for a graphical depiction of $b_1$ in Model 4.

Another simple transformation is to the mean center predictors that are being multiplied together prior to constructing the product. Such mean centering is often advocated when estimating models such as Equation 2, although the advantages of doing so are generally overstated (see, e.g., Cronbach, 1987; Kromrey & Foster-Johnson, 1998). To mean center a variable, the sample mean is subtracted from each measurement. Had this been done prior to calculating the product of ideology and perceived reality, the resulting model would be as presented in the model 5 column of Table 1. The coefficient for perceived reality is positive and statistically significant, $b_1 = 0.099$, $p = .034$ (see Figure 2 for a visual depiction). This represents the effect of perceived realism estimated at the sample mean political ideology (as a value of zero on political ideology corresponds to the sample mean after mean centering). If $M$ is a discrete variable, the regression coefficient for $X$ after mean centering can be interpreted as the weighted average conditional effect of $X$ across the values of $M$, with the weighting based on the relative frequency of the values of $M$. This is *not* the same thing as a main effect in ANOVA, which is an *unweighted* average conditional effect.

Another approach to estimating models with interactions is to standardize the variables being used to produce the interaction term prior to multiplying them together. This is similar to mean centering, in that a score of zero on variables that are standardized corresponds to the sample mean when unstandardized. However, the unit of measurement changes from the original metric to standard deviations. That is, a score of 1 corresponds to a response one standard deviation above the mean, and two cases that differ by 1 differ by a standard deviation rather than one point on the original metric. The final column in Table 1 (model 6) represents the coefficients from the model after first standardizing ideology and perceived reality prior to computing their product. Notice that the $t$ statistic and $p$-value for perceived reality have not changed compared to mean centering. This is because the hypothesis test still tests, as it did in the model in column 5, the null hypothesis that the conditional association between perceived reality and support for social spending is zero at the sample mean ideology. But the size of the coefficient for perceived reality has changed, reflecting the fact that two cases that differ by one unit on the original metric of measurement will differ by some different amount when the metric is changed to standard deviations. In this case, $b_1$ is smaller compared to when mean centering is used because the standard deviation for perceived reality is smaller than one.

Two additional points need to be made. First, one might argue that if researchers followed the practice of reporting standardized coefficients or expressing effects in terms of explained variance rather than reporting unstandardized regression coefficients, the differences reported above would vanish. That is, given that researchers’ decisions about measurement are always somewhat arbitrary, would standardizing variables prior to analysis not make the effect of such arbitrary decisions on statistical results go away? Unfortunately, this is simply not true, as is illustrated in Table 1. The standardized regression coefficients and variance explained for perceived reality also change when one of the linear transformations described here is applied to the data. The unit of measurement does not change the basic fact that any measure of a variable’s effect, in the presence of an interaction involving that variable as in Equation 2, is conditional on the other variable in the interaction equaling zero. Furthermore, the outcome of a test of the null hypothesis that a variable's effect is equal to zero is mathematically equivalent regardless of whether the effect is expressed as an unstandardized regression coefficient, a standardized coefficient, a semipartial or partial correlation, or an increase in $R^2$ when a variable is added to the model.

Second, Table 1 shows that the arbitrary choices of scaling described above have no effect on the test for the interaction. Whether one mean centers, standardizes, or adds or subtracts a constant from one or both of the variables involved in the interaction, the result of the null hypothesis test of no interaction (i.e., its $t$ statistic and $p$-value) in a model such as Equation 2 is unchanged, as are measures of effect size for the interaction. This might come as a surprise to those who advocate centering or standardizing in regression models on the grounds that it reduces multicollinearity and the problems multicollinearity produces. Although the regression coefficients, $t$ statistics, $p$-values, and measures of effect size are changed for those variables involved in the interaction, this has nothing to do with reducing multicollinearity. These differences are attributable to the effects of rescaling a variable in such a way that the "zero" point is changed vis-à-vis the original scale of measurement. The need to center (or standardize) $X$ and $M$ in a regression model including $XM$ as a predictor is a myth that doggedly persists in spite of having been repeatedly debunked (Cronbach, 1987; Echambi & Hess, 2007; Friedrich, 1982; Hayes, 2005, pp. 465–467; Kromrey & Foster-Johnson, 1998; Shieh, 2011).

That said, there are some advantages to mean centering or standardizing. First, mean centering $X$ and $M$ prior to computation of their product produces an estimate of the effect of $X$ (or $M$) on $Y$ that is conditioned on a value of $M$ (or $X$) that will always be within the range of the data (i.e., at the sample mean), unlike when mean centering is not done. Thus, by mean centering, the consequences of the kind of misinterpretations we describe here will be much less severe when they occur. Second, in complex models with *several* interactions involving the same variable, multicollinearity can in rare circumstances produce computational difficulties in some statistical programs. Mean centering $X$ and $M$ prior to computing their product does indeed reduce this "nonessential" multicollinearity and thereby eases the computation problems that sometimes though rarely arise. In the majority of applications of moderated multiple regression in the field of communication, however, mean centering or standardization is a choice one can make, to do or not to do, rather than a requirement.[3]

## CONCLUSION

In this editorial primer, we have attempted to clarify the interpretation of parameter estimates and hypothesis tests from a regression model with interactions. When $X$, $M$, and $XM$ all coexist in a regression model, the coefficients and tests of significance for $X$ and $M$ are estimated and tests of *conditional* effects or *simple* effects, not main effects as they are in ANOVA and sometimes interpreted as in the communication literature. They estimate the effect of one variable

[3]We recommend that dichotomous variables not be standardized under any circumstances, for the regression coefficient for a standardized dichotomy has no sensible interpretation and is determined by the distribution of the sample across the two groups coded with the dichotomous variable (see e.g., Cohen et al., 2003, p. 316). This applies to regression models without interactions, and to the interpretation of standardized regression coefficients for dichotomous predictors as well.

conditioned on the other equaling zero. They cannot be interpreted as partial regression coefficients in a model without an interaction, nor are they equivalent to main effects in ANOVA. A greater understanding of these subtle but very important principles will reduce the number of claims in the communication literature that are not substantiated by statistical evidence or that are based on statistical information from a linear model that is not pertinent to the claim being made.

## REFERENCES


Aguinis, H. (2003). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373–400.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426–443.

Cohen, J. (1978). Partialed products *are* interactions; partialed powers *are* curve components. *Psychological Bulletin*, *85*, 858–866.

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

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin*, *102*, 414–417.

Darlington, R. B. (1990). *Regression and linear models*. New York, NY: McGraw-Hill.

Echambi, R., & Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Research*, *26*, 438–445.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, *26*, 797–833.

Glynn, C. J., Huge, M. E., Reineke, J. B., Hardy, B. W., & Shanahan, J. (2007). When Oprah intervenes: Political correlates of daytime talk show viewing. *Journal of Broadcasting and Electronic Media*, *51*, 228–244.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, *41*, 924–936.

Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, *38*, 100–109.

Jaccard, J. J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. Thousand Oaks, CA: Sage Publications.

Kam, C. D., & Franzese, R. J. (2007). *Modeling and interpreting interactive hypotheses in regression analysis*. Ann Arbor, MI: University of Michigan Press.

Kromrey, J. D., & Foster-Johnson, L. (1998). Mean centering in moderated multiple regression: Much ado about nothing. *Educational and Psychological Measurement*, *58*, 42–67.

Shieh, G. (2011). Clarifying the role of mean centering in multicollinearity of interaction effects. *British Journal of Mathematical and Statistical Psychology*, *64*, 462–477.

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology*, *19*, 111–120.