Tests of Moderation Effects: Difference in Simple Slopes versus the Interaction Term

Cecil D. Robinson **Sara Tomek** **Randall E. Schumacker**
University of Alabama

This paper analyzes two methods for testing moderation effects in regression models that contain a continuous dependent variable, a continuous independent variable, and a dichotomized grouped moderator; a test of the interaction term in the full regression model and a test of the difference between the simple slopes. Typically, researchers test the significance of the interaction term. Based on mathematical equations and empirical examples, we argue that the test for the difference between the simple slopes should be utilized when researchers are interested in testing for moderation effects. By decomposing the test statistics for these two methods, we demonstrate that the test for the difference between the simple slopes has increased power and less Type II error, while retaining equivalent Type I error rates.

Moderated relationships in social science research exist when the relationship between two variables, X and Y, varies depending on the value of a third variable Z. This study examines a specific type of moderated relationship with a continuous dependent variable (Y), a continuous independent variable (X), and an independent dichotomous categorical variable (Z). Given these variables, a moderated relationship exists if the relationship between X and Y is different for both levels of Z. This can be estimated with an interaction term using the following regression equation (Aiken & West, 1991; Jaccard & Turrisi, 2003).

$$Y = b_1X + b_2Z + b_3XZ + b_0 \,; \tag{1}$$

where: Y = continuous dependent variable,
X = continuous independent variable,
Z = dichotomous independent variable,
XZ is the interaction term calculated as X multiplied by Z,
b_0 is the intercept,
b_1 is the effect of X on Y,
b_2 is the effect of Z on Y, and
b_3 is the effect of XZ on Y.

To understand how the interaction term XZ tests for a moderated relationship, consider Equation 1. With two values of Z, we can use dummy coding such that Group 1 is 0 (Z = 0) and Group 2 is 1 (Z = 1). By substituting 0 for Z for Group 1, Equation 1 becomes:

$$Y = b_1X + b_2(0) + b_3X(0) + b_0 \tag{2a}$$

By removing the 0 terms, this can be further reduced to

$$Y = b_1X + b_0 \tag{2b}$$

Next, we will substitute 1 for Z for Group 2; therefore, Equation 1 becomes

$$Y = b_1X + b_2(1) + b_3X(1) + b_0 \tag{3a}$$

After grouping terms, this is represented as

$$Y = (b_1 + b_3)X + (b_2 + b_0) \tag{3b}$$

Within Equations 2b and 3b, the slopes of X for Groups 1 and 2 are b_1 and $(b_1 + b_3)$, respectively. The difference in the two slopes is $(b_1 + b_3) - b_1 = b_3$; the coefficient of the interaction term XZ in Equation 1. Therefore, if the results of the regression model in Equation 1 indicate that the interaction term, b_3, is significant, then the slopes of X for Groups 1 and 2 (Z) are statistically different from each other, indicating a statistically significant moderated relationship between X and Y.

Once a significant moderated relationship has been identified, a researcher can examine the strength of relationships between X and Y within the individual groups of Z. To accomplish this, data are separated by group and then individual regression equations are estimated with X regressed on Y. These are termed the simple slopes. In our example with two levels of Z, the simple slopes require two regression models, one for each group, and are estimated with the following:

$$\text{For Group 1:} \quad Y = b_{1,Z=0}\,X + b_{0,Z=0} \tag{4}$$

$$\text{For Group 2:} \quad Y = b_{1,Z=1}\,X + b_{0,Z=1} \tag{5}$$

These two models will estimate strength of the individual betas for X as well as the overall fit of the models for each group. Analysis of simple slopes provides additional information not produced within the full interaction term model (Equation 1). While the simple slopes models provide the researcher with additional information, the estimates of the betas in the simple slopes models are equivalent to the estimates of the betas in the full interaction term model as outlined in Table 1. The simple slopes models do not automatically test the significance of the difference in the simple slopes ($b_{1,Z=1} - b_{1,Z=0}$; simple slopes difference). The full interaction model will provide this test in the estimate, b_3, which is justification to researchers as to why an interaction model must establish significance before examining the simple slopes.

Table 1. Equivalent Estimates Between Simple Slopes Models and the Interaction Model

Interaction Model		Equivalent Simple Slopes Models
Term	Estimate	Estimate
Intercept	b_0	$b_{0,Z=0}$
X	b_1	$b_{1,Z=0}$
Z	b_2	$(b_{0,Z=1} - b_{0,Z=0})$
XZ	b_3	$(b_{1,Z=1} - b_{1,Z=0})$

Within this study; however, we present the method for testing the significance of the difference in the simple slopes. Further, we argue that the test of significance for the difference in the simple slopes is more powerful than the test of significance for the interaction term, b_3 (Equation 1). Based on this logic, if researchers are interested exclusively in testing the moderated relationship, then we recommend testing the difference in the simple slopes in lieu of testing the significance of the interaction term, b_3 (Equation 1).

In this study, two empirical examples will be presented to demonstrate the differences between the two approaches and to highlight the practical implications of these differences. Within these examples, the differences between the two methods will be analyzed mathematically. The differences between these two methods can be found by decomposing how standard error is computed for each method. Based on our findings, we argue that (1) the use of the simple slopes models alone provides more information to the researcher than testing the interaction term first; (2) testing for the difference in the simple slopes has more statistical power than the test for the interaction term by reducing the threat of Type II error; and (3) testing the significance of the difference in simple slopes does not increase Type I error.

Testing for Significance of the Difference in Simple Slopes

To test whether the individual slopes of the simple slopes models are different, a *t*-value for the difference in the two slopes can be computed using the following equation:

$$t = \frac{b_{\text{diff}}}{SE_{\text{pooled}}} \tag{6}$$

The difference in the two slopes from Equations 2b and 3b is:

$$b_{\text{diff}} = b_1 - b_2 \tag{7}$$

The pooled standard error from Equations 2b and 3b is:

$$SEpooled = \sqrt{\frac{n_1 SE_1^2 + n_2 SE_2^2}{n_1 + n_2 - 2}} \tag{8}$$

Substituting b_{diff} and SE_{pooled} in Equation 4 results in the following *t*-test:

$$t = \frac{b1 - b2}{\sqrt{\frac{n_1 SE_1^2 + n_2 SE_2^2}{n_1 + n_2 - 2}}} \tag{9}$$

Table 2. Full Regression and Simple Slopes Models of Academic Self-Efficacy (ASE), Ethnicity, and Academic Achievement.

Full Regression Model (n = 209)				
	b	SE	t	p
Intercept	2.957	.069	42.590	.000
ASE	.031	.006	4.956	.000
Ethnicity	-.098	.098	-.998	.320
Ethnicity x ASE	-.013	.00932	-1.403	.162

Simple Slopes Models				
	b	SE	t	p
African-American Students (n = 105)				
Intercept	2.859	.067	42.930	.000
ASE	.018	.00669	2.687	.008
Caucasian Students (n = 104)				
Intercept	2.957	.072	41.113	.000
ASE	.031	.00649	4.784	.000
Simple Slopes Difference				
	-.013	.00659	-1.973	.0498

Once one has computed the t-value for the difference of the simple slopes, then it is straightforward to determine the p-value for the test statistic with degrees of freedom ($n_1 + n_2 - 2$).

Empirical Examples Comparing the Differences in Simple Slopes and the Interaction Term

Example 1. The first example was taken from Robinson and Schumacker (2009). This study examined the effects of centering on multicollinearity for interaction terms in regression models (Equation 1) with a continuous dependent variable (Y), one continuous independent variable (X), and one dichotomous independent variable (Z). The data from this study examined the relationship between ethnicity and academic self-efficacy (ASE) on academic achievement. Estimates from the full regression model examining the main effects and interaction between ethnicity and ASE (XZ) on academic achievement are presented in Table 2. Although ASE is significantly related to academic achievement ($p < .001$), ethnicity was not related to academic achievement ($p = .320$) and the interaction term (ethnicity x motivation) indicated that the relationship between motivation and academic achievement did not vary by ethnicity ($p = .162$).

The estimated simple slopes models (Equations 4 and 5) examining the relationship between motivation and academic achievement based on ethnicity are also presented in Table 2. The two models indicate that ASE is significantly related to academic achievement for both African-American ($p = .008$) and Caucasian ($p < .001$) students. To test for differences between the two groups in their relationship between ASE and academic achievement, a test for the differences between the simple slopes was estimated. First, the difference in the betas was calculated, $b_{diff} = b_{African\text{-}American} - b_{Caucasian} = .018 - .031 = -.013$ (Equation 7). Next, the pooled standard error was calculated, $SE_{pooled} = \sqrt{\frac{n_1 SE_1^2 + n_2 SE_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{105(.00669)+104(.00649)}{105+104-2}} = .00659$ (Equation 8). Third, the b_{diff} and SE_{pooled} were used to compute the t-test, $t = \frac{b_{diff}}{SE_{pooled}} = \frac{-.013}{.00659} = -1.973$ (Equation 9). Finally, the t-distribution was used to determine the p-value for the test, $t\,(207) = -1.973$, $p = .0498$. This secondary analysis indicated a statistically significant difference in the simple slopes for African-American and Caucasian students, or that ethnicity was a statistically significant moderator of ASE and academic achievement.

What is interesting is that the analysis of the interaction term and the test for the differences in the simple slopes would indicate two different interpretations of the same dataset; the interaction term was not significant, yet significant differences in the simple slopes were found. If a researcher used the interaction term XZ (Equation 1) as a test of moderation, then the researcher would infer that ethnicity did not moderate the relationship between ASE and academic achievement. However, if the researcher used the difference in simple slopes, then the researcher would infer that ethnicity did moderate the relationship between ASE and academic achievement. If the researcher followed the practice of testing the interaction term first, then the simple slopes would not have been estimated. In this example, the choice of analytic approach is the difference between reporting a significant moderation relationship and reporting a non-significant moderation relationship.

Example 2. The second example is a dataset (home7dat.txt) taken from Aiken's Multiple Regression Course at Arizona State University (http://www.public.asu.edu/~atlsa/PSY531/). This dataset was chosen because in Aiken's course, it is used to teach about testing interactions between a dichotomous categorical variable and a continuous variable. These data are used to examine whether grouping individuals by high risk or low risk of breast cancer (group) moderated the relationship between risk of breast cancer (risk) and intention to obtain screening (intention), which used Model 1 to estimate this effect.

The full model estimates are presented in Table 3. Results indicated that group ($p < .001$) and risk ($p < .001$) were significantly related to intention, and that the interaction between risk and group was significantly related to intention ($p < .001$).

The estimated simple slopes models (Equations 4 and 5) examining the relationship between risk and intention based on high and low risk groups are also presented in Table 3. Individually, the two models indicate that risk was significantly related to intention for both high ($p < .001$) and low ($p < .001$) risk groups, albeit in different directions. To test whether the relationship between risk and intention was different for the two risk groups, a test for difference in simple slopes was calculated; $b_{diff} = b_{Low\ Risk} - b_{High\ Risk} = .908 - (-.951) = -1.859$, $SE_{pooled} = .155$, $t = 1.859 / .155 = -11.991$. This analysis indicated a significant difference in the slopes of the high and low risk groups, $t\ (184) = -11.991$, $p < .001$, or that group significantly moderated the relationship between risk and intention.

Unlike Example 1, both analyses of Example 2 led the researcher to make the same inferences about the moderated relationship between risk and intention. There is; however, a difference in the resulting *t*-statistics for the two hypothesis tests; -8.55 for the test of the interaction term (Equation 1) and -11.991 for the difference between the simple slopes (Equation 7). Although there is no difference in interpretation in this example, the test of difference in simple slopes again demonstrates greater power as compared to the test of the interaction term.

Standard Error Differences between the Difference in Simple Slopes and the Interaction Term

The two empirical examples demonstrated differences between the test for the difference in simple slopes and the test for the interaction term in identifying moderation effects. Further examination of the test values used for significance testing indicates differences in the standard errors (SEs) of the two approaches (see Table 4). The beta of the interaction term and the difference between the simple slopes are mathematically identical and; therefore, also identical in the Robinson and Schumacker (2009) dataset (-.013) and the Aiken (http://www.public.asu.edu/~atlsa/PSY531/) dataset (1.859). However, there are differences in the SEs. Specifically, the SEs for the difference in simple slopes are consistently smaller than the SEs for the interaction term; 28.6% less in the Aiken dataset and 29.3% less in the Robinson and Schumacker dataset. The result of a smaller SE in the denominator when computing *t* is a greater *t*-statistic.

Table 3. Full Regression and Simple Slopes Models of Risk, Risk Grouping, and Intention to Get Cancer Screening

Full Regression Model (n = 186)					Simple Slopes Models				
	b	*SE*	*t*	*p*		*b*	*SE*	*T*	*p*
Intercept	7.127	.224	31.774	.000	Low Risk Group (n = 95)				
Group	-2.317	.329	-7.044	.000	Intercept	7.127	.213	33.485	.000
Risk	.908	.149	6.103	.000	Risk	.908	.141	6.432	.000
Group x Risk	-1.860	.217	-8.555	.000	High Risk Group (n = 91)				
					Intercept	4.810	.253	19.014	.000
					Risk	-.951	.167	-5.712	.000
					Simple Slopes Difference				
						1.859	.155	11.994	.000

Table 4. Comparison of Values Used to Test for Significance of Interaction Term and Difference in Simple Slopes

	Aiken Dataset				Robinson and Schumacker Dataset			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Interaction Term	-1.859	.217	-8.555	.000	-.013	.00932	-1.403	.162
Simple Slopes Difference	-1.859	.155	-11.991	.000	-.013	.00659	-1.973	.0498
SE Difference		.062 (28.6%)				.00273 (29.3%)		

In an attempt to understand why these differences occurred, we decomposed the SE of the interaction term. The following formula for the SE of the interaction term can also be used to compute the SEs for the two simple slopes. The SEs of the simple slopes are then pooled when calculating the *t*-test for difference in simple slopes (Equation 7). By decomposing the SE, differences between the two hypothesis tests can be identified. The SE for the significance test of the interaction term in Equation 1 is computed using the following.

$$\mathrm{SE} = \sqrt{\frac{\mathrm{MSE}}{\mathrm{SS} * \mathrm{TOL}}} \tag{10}$$

Mean squared error (MSE) is computed using the following.

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2 \tag{11}$$

Sum of squares (SS) is computed using the following.

$$\mathrm{SS} = \sum (XZ)^2 - \frac{(\sum XZ)^2}{n} \tag{12}$$

Tolerance is computed using the following.

$$\mathrm{TOL} = 1 - R^2_{XZ|X,Z} \tag{13}$$

Comparison of the Mean Squared Error for the Difference in Simple Slopes and the Interaction Term

Mean squared error (MSE) is the residual error not explained by the regression model. It is measured as the sum of the squared difference between the estimated and the true score of Y. Since the estimate of the beta of the interaction term (Equation 1) is equivalent to the difference in the betas of the simple slope models (Equations 4 and 5), the relationship between the interaction term and the dependent variable is identical to the combined relationship of the betas in the simple slopes models and the dependent variable. Therefore, the residual error, MSE (Equation 11), for these two approaches should also be equivalent. Indeed, the weighted average of the MSEs for the simple slopes models (Equations 4 and 5) are equivalent to the MSE of the full model (Equation 1) (see Table 5). Because the MSEs are equivalent, we further examined Equation 12 and as to whether the source of differences in the SEs was in the sum of squares.

Table 5. MSE for Full Model, Simple Slopes Models, and Weighted Average of Simple Slopes Models

Aiken Dataset		Robinson and Schumacker Dataset	
	MSE		*MSE*
Interaction Term	2.591	Interaction Term	.498
Simple Slopes Average	2.593	Simple Slopes Average	.498
Low Risk	2.333	Caucasians	.534
High Risk	2.864	African-Americans	.462

Comparison of the Sum of Squares for the Difference in Simple Slopes and the Interaction Term

The sum of squares (SS) is a measure of dispersion. The lower the SS, the more homogeneous the dataset and the larger the SS, the more heterogeneous the dataset. Table 6 presents the computation of the SS (Equation 12) used in the estimation of the SE in both the simple slopes models (Equations 4 and 5) and the full model with the interaction term (Equation 1) for Aiken's (http://www.public.asu.edu/~atlsa/PSY531/) dataset. Looking at the results, the sum of X^2 (Step 1) and the square of the sum of X (Step 2) are identical to the simple slopes for the high risk group and the risk x group interaction term. This is true due to the dummy coding of the group variable, 0 (low risk) and 1 (high risk). When group is multiplied by risk, all that remains are the values when group equals 1 (high risk); all low risk X values are equal to 0. However, a difference does occur when dividing the square of the sum of X by n (Step 3). The model including the interaction term (Equation 1) contains the full sample (n = 186), whereas the simple slopes model for high risk group only contains those within the high risk group (n = 91). This produces the smaller term for the full model with the interaction term (Equation 1) than the simple slope model for the group where Z=1 (high risk group; Equation 5) in Step 3 (51.916 versus 106.684). The smaller term in Step 3 produces a larger SS for the interaction term in Step 4 when subtracted from the sum of X^2 from Step 2 (SS = 158.001) as compared to the SS (Step 4) of the high risk group, Z = 1 (SS = 103.232). An identical result is found in the Robinson and Schumacker (2009) dataset. The SS (Step 4) for the interaction term in the full model is larger (SS = 10378.251) than the SS of the simple slope model for African-Americans, Z = 1 (SS = 10332.569).

This empirical analysis highlights that the SS used for the interaction model will always be greater than the SS used for the simple slope for the grouping variable, Z, dummy coded as 1 (high risk group for Aiken (http://www.public.asu.edu/~atlsa/PSY531/); African-American for Robinson & Shumacker, 2009). This is true because the sum of Xs for the interaction term in Steps 1 and 2 will always be identical to the sum of Xs in Steps 1 and 2 for the simple slopes model. However, the SS for the interaction term is a function of the full sample whereas the dummy coded grouping models will only include subset of the full sample. The SS for the interaction term model will always include a larger n as a divisor; therefore, the interaction SS will always be larger than either of the simple slopes' SS in Step 4.

Although the SS for the interaction term will always be larger than the SS for the simple slope for the grouping variable, Z = 1, the same cannot be said for the SS for the grouping variable, Z, dummy coded as 0. In the Aiken (http://www.public.asu.edu/~atlsa/PSY531/) dataset, the SS for the simple slope of the low risk group (SS = 117.044) was less than the SS of the for the interaction term (SS = 158.001). However, in the Robinson and Schumacker (2009) dataset, the SS for the simple slope of Caucasian student group (SS = 12649.818) was greater than the SS of the interaction term (SS = 10378.251).

The analyses of the SSs do not conclusively identify the source of the lower SEs for the difference in simple slopes compared to the interaction term. Therefore, tolerance was examined as this is the last term in Equation 10 and must account for the differences in the test values for the two procedures.

Comparison of the Tolerance for the Difference in Simple Slopes and the Interaction Term

Tolerance (TOL) is the proportion of variance for an independent variable that is not accounted for by other independent variables in a regression model. TOL is measured as one minus the squared multiple correlation of the target variable regressed on all other independent variables. The range of TOL is 0 to 1, with smaller values indicating a larger concern for multicollinearity.

Table 6. Decomposition of Sums of Squares for Simple Slopes and Interaction Models

Components for Sum of Squares	Aiken Dataset			Robinson and Schumacker Dataset		
	Simple Slopes Models		Full Model	Simple Slopes Models		Full Model
	Risk		Risk x Group Interaction	ASE		ASE x Ethnicity Interaction
	Low	High		Caucasian	African-American	
n	95	91	186	104	105	209
Step 1: $\sum X^2$	218.172	209.917	209.917	12742.070	10424.372	10434.372
Step 2: $(\sum X)^2$	9607.16	9708.277	9708.277	9594.203	9639.312	9639.312
Step 3: $\frac{(\sum X)^2}{n}$	101.128	106.684	51.916	92.252	91.803	46.121
Step 4: $\sum X^2 - \frac{(\sum X)^2}{n}$	117.044	103.232	158.001	12649.818	10332.569	10378.251

Table 7. Effect of the Tolerance on the Standard Error

	Aiken Dataset			Robinson and Schumacker Dataset		
	Simple Slopes Models		Full Model	Simple Slopes Models		Full Model
	Risk		Risk x Group Interaction	ASE		ASE x Ethnicity Interaction
	Low	High		Caucasians	African-Americans	
n	95	91	186	104	105	209
$R^2_{XZ\|X,Z}$	--	--	.649	--	--	.447
$R^2_{X\|-}$	0	0	--	0	0	--
TOL	1	1	.351	1	1	.553
MSE	2.333	2.864	2.591	.534	.462	.498
SS	117.044	103.232	158.001	12649.818	10332.569	10378.251
*SS*TOL*	117.044	103.232	55.458	12649.818	10332.569	5687.282
$SE = \sqrt{\frac{MSE}{SS * TOL}}$	.141	.167	.216	.00649	.00669	.00932
SE_{pooled}	.155			.00659		

Note: ASE = Academic Self-Efficacy; TOL = Tolerance; MSE = Mean Squared Error;
SS = Sum of Squares; SE = Standard Error

The TOL values for the two simple slopes terms (Equations 4 and 5) and the interaction term (Equation1) for the two example datasets are presented in Table 7. For the Aiken (http://www.public.asu.edu/~atlsa/PSY531/) dataset, the TOL = .351 for the interaction term XZ (Equation 1). TOL = 1 for the simple slopes (Equations 4 and 5). In the Robinson and Schumacker (2009) dataset, the TOL = .553 for the interaction term and TOL = 1 for the simple slopes.

While the TOL of the interaction term (XZ) is equal to $1 - R^2_{XZ|X,Z}$, the TOL of the simple slopes models is always equal to 1. The presence of tolerance in the test statistic for the interaction model within the full regression model (Equation 1) is to statistically control for multicollinearity between the variables. However, the manual process of grouping the values of X by Z within the simple slopes models (Equations 4 and 5) eliminates any potential multicollinearity because there are no other terms within the regression models for the X terms with whom to correlate. Since there are no other terms in the simple slopes model, the squared multiple correlation of X with no other independent variables, $R^2_{X|-}$, is 0, and the TOL(X) = 1 – 0 = 1. To understand the impact of the TOL on the SE, we revisit the calculations for the two hypothesis tests.

Revisiting the Standard Error Differences between the Difference in Simple Slopes and the Interaction Term

Table 7 contains the three components of the SE that were previously discussed in this study: MSE, SS, and TOL. Let us begin our discussion with the Aiken (http://www.public.asu.edu/~atlsa/PSY531/) dataset. In the denominator of the SE term (Equation 10), the TOL of the interaction term is multiplied by the SS. For this example, SS x TOL resulted in the value 54.826. When compared to the denominator of the SEs for the simple slopes models (117.044 and 103.232), the denominator of the SE for the interaction term is smaller. A smaller SE denominator for the interaction term results in a larger SE for the interaction term. The SE for the interaction term was equal to SE = .216, while the SE_{pooled} used to test for the difference in simples slopes was equal to SE_{pooled} = .155. An identical comparison can be made for the Robinson and Schumacker (2009) dataset. The denominator of the SE for the interaction term is smaller

(5687.282) than either of the denominators for the simple slopes models (12649.818 and 10332.569, respectively). The resulting SE for the interaction term (SE = .00932) is larger than the SE_{pooled} used to test for the difference in simple slopes (SE_{pooled} = .00659).

We argue that the multicollinearity present within the interaction model (Equation 1) reduces the power to detect statistical significance in the interaction term. As an example, within the Aiken (http://www.public.asu.edu/~atlsa/PSY531/) dataset, the tolerance must increase by 194.6% – from .351 to .683 – in order for the SE of the interaction term to equal the SE_{pooled} of the simple slopes. Stated another way, the $R^2_{XZ|X,Z}$ must decrease from .649 to .317. This decrease is problematic on at least two accounts. First, reducing the relationship between the interaction term XZ and X and Z appears to defeat the purpose of testing for a moderating effect (also note that centering the independent variables will not reduce this relationship). Second, this reduction in the relationship does not seem feasible as one would generally expect $R^2_{XZ|X,Z}$ to be high whether the interaction term is significant or not, as XZ is a composite of its two variables X and Z.

The denominators of the SE terms in the Robinson and Schumacker (2009) dataset were also compared (Table 7), with similar results detected. In order for the SE of the interaction term to equal the SE_{pooled} of the simple slopes, the tolerance of the interaction term must increase 199.8%, from .553 to 1.105; or the $R^2_{XZ|X,Z}$ must decrease from .452 to -.105. There are similar proportional increases in tolerance that are required (194.6% versus 199.8%). However, the results from the Robinson and Shumacker dataset are more problematic, as it is not mathematically possible for $R^2_{XZ|X,Z}$ to be less than zero and the tolerance has an upper limit of one. This demonstrates an instance where the test for the interaction term will *always* be less powerful than the test for the simple slopes difference, as it is impossible for the $R^2_{XZ|X,Z}$ to be less than zero; therefore, the test statistic for the interaction term must be smaller than the test statistic for the difference in simple slopes.

Implications

Based on the equations and examples presented above, we argue that researchers investigating moderator variables should begin with an examination of the simple slopes rather than relying on a significant interaction term. In both examples, the test for the difference in simple slopes was more powerful than the test of the interaction terms. Even when the interaction term was significant (as in Aiken's dataset), the test for the difference in simple slopes demonstrated more power (it had a larger *t*-value). Further, testing the significance of the difference in simple slopes involves only one hypothesis test (although we need to estimate 2 models, Equations 4 and 5). This is identical to the test for the interaction term, where one hypothesis test (estimation of a single model, Equation 1) is performed. Therefore, testing for the difference in simple slopes does not increase the Type I error, as this is based on the number of hypothesis tests performed. Perhaps the only downside to this approach is that the researcher will need to manually compute the *t*-value for the difference in simple slopes, as this is not automatically calculated in SPSS or SAS.

Discussion

While pondering the implications of these results, one question that was considered was why, historically, the interaction term has been preferred to the difference in simple slopes? This question is important for two reasons. First, the formulas presented in this study are not new, yet we were unable to find any previous literature that compared (either conceptually or empirically) the difference of simple slopes and the interaction term in a multiple regression context. Second, little research can be found that discusses the mathematics behind the test for the difference in simple slopes. Few textbooks devote time to discussing methods for testing differences between separate regression equation coefficients (note: our equations were verified using Kleinbaum and Kupper's, 1978 textbook, which is several decades old).

Parsimony may be one potential reason that the two approaches have not been directly compared. Prior to computers, test statistics were computed by hand. Hierarchically adding an interaction term after computing the additive model requires one additional set of computations. Only if the researcher finds a significant interaction term would they then consider computing the two additional simple slopes models. However, testing for the difference in simple slopes requires computing two simple slopes models prior to detecting a significant difference between them. It is not difficult to imagine the frustration a researcher would experience when their time spent calculating these models did not result in a significant difference.

Today, there would be no such concern provided the ease of computation that statistical packages allow (although the frustration of failing to reject the null still remains).

Another reason these two approaches may not have been directly compared is that the two approaches ask different research questions. The interaction term tests whether the composite of XZ accounts for a significant amount of variance in Y beyond the additive effects of X and Z. The differences in the simple slopes tests whether there is a different relationship between X and Y for each group Z. Although the two approaches represent different research questions, both approaches attempt to statistically answer whether there is a significant moderating relationship among the variables. Conceptually, we believe that the difference in simple slopes is a more direct test of moderation. Empirically, the test for the difference in the simple slopes has more power, which reduces the potential for Type II Error and simultaneously does not increase Type I Error.

Future Research

Based on our empirical findings, we believe that there needs to be a review of published and unpublished studies to inform the severity of Type II Error in testing the interaction term prior to estimating the simple slopes. Significant effects may have been missed in previous studies (e.g., this result would have occurred in the Robinson and Schumacker, 2009 dataset). Second, simulation studies will be required in order to determine if there is a single "critical point" (if it exists) for TOL such that the SEs of the interaction term and difference in the simple slopes are equivalent. Additionally, simulation studies will help to determine whether sample sizes, unbalanced designs, and/or variance within the moderator groups affect the "critical point" of TOL. Finally, this paper examined one specific type of moderated relationship, that which has a single continuous dependent variable, a single continuous independent variable, and a single dichotomous categorical independent variable. Additional calculations are needed in order to determine if our findings can be replicated to other types of moderated relationships (e.g., a categorical independent variable with more than two levels), or if these finding are specific to this type of relationship.

Conclusion

The current study critically analyzed the differences and similarities between two hypothesis tests: the test for the difference in simple slopes and the test for the interaction term. Our results indicate that empirically, in the test for the difference in simple slopes, the denominator of the SE will be larger, the standard error will be smaller, and the resulting test statistic will be larger. As a result, the test for the difference in simple slopes will have more power (decrease Type II error), yet retain an equivalent level of Type I error. The test for the difference in simple slopes also makes more conceptual sense, as it is a direct test for the moderating effect of the grouping variable Z. Overall, we recommend the use of the test for the difference in simple slopes directly when researchers are interested in testing the moderating relationship of a dichotomous grouping variable, Z, on the relationship between two continuous variables, X and Y.

References

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

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Sage Publications: Thousand Oaks, CA.

Kleinbaum, D. G., & Kupper, L. L. (1978). *Applied regression analysis and other multivariable methods*. North Scituate, MA: Duxbury Press.

Robinson, C., & Schumacker, R. (2009). Interaction effects: Centering, variance inflation factor, and interpretation issues, *Multiple Linear Regression Viewpoints, 35*(1), 6–11.

Send correspondence to: Cecil D. Robinson
University of Alabama
Email: crobinso@bamaed.ua.edu