White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/114588/

Version: Accepted Version

Article:
Dawson, J. (2014) Moderation in management research: What, why, when and how. Journal of Business and Psychology, 29 (1). pp. 1-19. ISSN: 0889-3268

https://doi.org/10.1007/s10869-013-9308-7

© 2013 Springer Science+Business Media New York. This is an author produced version of a paper subsequently published in Journal of Business and Psychology. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Moderation in management research: What, why, when and how

Jeremy F. Dawson

University of Sheffield, United Kingdom

Citing this article:

Dawson, J. F. (2014). Moderation in management research: What, why, when and how. *Journal of Business and Psychology, 29,* 1-19.

The final publication is available at Springer via http://dx.doi.org/10.1007/s10869-013-9308-7

Abstract

Many theories in management, psychology and other disciplines rely on moderating variables: those which affect the strength or nature of the relationship between two other variables. Despite the near-ubiquitous nature of such effects, the methods for testing and interpreting them are not always well understood. This article introduces the concept of moderation and describes how moderator effects are tested and interpreted for a series of model types, beginning with straightforward two-way interactions with Normal outcomes, moving to three-way and curvilinear interactions, and then to models with non-Normal outcomes including binary logistic regression and Poisson regression. In particular, methods of interpreting and probing these latter model types, such as simple slope analysis and slope difference tests, are described. It then gives answers to twelve frequently asked questions about testing and interpreting moderator effects.

Moderation in management research: What, why, when and how

This paper is the eighth in this journal's *Method Corner* series. Previous articles have included topics encountered by many researchers such as tests of mediation, longitudinal data, polynomial regression, relative importance of predictors in regression models, common method bias, construction of higher order constructs, and most recently combining structural equation modeling with meta-analysis (c.f., Johnson, Rosen, & Chang, 2011; Landis, 2013). The present article is designed to complement these valuable articles by explaining many of the issues surrounding one of the most common types of statistical model found in the management and organizational literature: moderation, or interaction effects.

Life is rarely straightforward. We may believe that exercising will help us to lose weight, or that earning more money will enable us to be happier, but these effects are likely to occur at different rates for different people – and in the latter example might even be reversed for some. Management research, like many other disciplines, is replete with theories suggesting that the relationship between two variables is dependent on a third variable; for example, according to goal-setting theory (Locke, Shaw, Saari & Latham, 1981), the setting of difficult goals at work is likely to have a more positive effect on performance for employees who have a higher level of task ability; while the Categorization-Elaboration Model of work group diversity (van Knippenberg, De Dreu & Homan, 2004) predicts that the effect of diversity on the elaboration of information within a group will depend on group members' affective and evaluative reactions to social categorization processes.

Many more examples abound; almost any issue of a journal containing quantitative research will include at least one article which tests what is known as moderation. In general

terms, a *moderator* is any variable that affects the association between two or more other variables; *moderation* is the effect the moderator has on this association. In this article I first explain how moderators work in statistical terms, and describe how they should be tested and interpreted with different types of data. I then provide answers to twelve frequently asked questions about moderation.

Testing and interpreting moderation in ordinary least squares regression models

Moderation in statistical terms

The simplest form of moderation is where a relationship between an independent variable, X, and a dependent variable, Y, changes according to the value of a moderator variable, Z. A straightforward test of a linear relationship between X and Y would be given by the regression equation of Y on X:

$$Y = b_0 + b_1X + \varepsilon \quad [1]$$

where b_0 is the intercept (the expected value of Y when X = 0), b_1 is the coefficient of X (the expected change in Y corresponding to a change of one unit in X), and ε is the residual (error term). The coefficient b_1 can be tested for statistical significance (i.e., whether there is evidence of a non-zero relationship between X and Y) by comparing the ratio of b_1 to its standard error with a known distribution (specifically, in this case, a *t*-distribution with $n - 2$ degrees of freedom, where *n* is the sample size).

For moderation, this is expanded to include not only the moderator variable, Z, but also the *interaction term* XZ created by multiplying X and Z together. This is called a *two-way interaction*, as it involves two variables (one independent variable and one moderator):

$$Y = b_0 + b_1X + b_2Z + b_3XZ + \varepsilon \quad [2]$$

This interaction term is at the heart of testing moderation. If (and only if) this term is significant – tested by comparing the ratio b_3 to its standard error with a known distribution – can we say that Z is a statistically significant moderator of the linear relationship between X and Y. The coefficients b_1 and b_2 determine whether there is any main effect of X or Z respectively, independent of the other, but it is only b_3 that determines whether we observe moderation.

Testing for two-way interactions

Due to the way moderation is defined statistically, testing for a two-way interaction is straightforward. It simply involves an ordinary least squares (OLS) regression in which the dependent variable, Y, is regressed on the interaction term XZ and the main effects X and Z.[1] The inclusion of the main effects is essential; without this, the regression equation (equation [2] above) would be incomplete, and the results cannot be interpreted. It is also possible to include any control variables (covariates) as required.

The first step, therefore, is to ensure the interaction term can be included. In some software (e.g. R) this can be done automatically within the procedure, without needing to create a new variable first. However, with other software (e.g. SPSS) it is not possible to do this within the standard regression procedure, and so a new variable should be calculated. Example syntax for such a calculation is shown within the appendix.

An important decision to make is whether to use the variables X and Z in their raw form, or to mean-center (or *z*-standardize) them before starting the process. In the vast majority of cases, this makes no difference to the detection of moderator effects; however, each method

[1] This test of moderation involves the same assumptions as does any "ordinary least squares" (OLS) regression analysis – i.e. residuals are independent and Normally distributed, and their variance is not related to predictors – and for most of this article I will assume this to be the case without further comment; I will deal separately with non-Normal outcomes later.

confers certain advantages in the interpretation of results. A fuller discussion in the merits of centering and *z*-standardizing variables can be found later in the article; I will assume, for now, that all continuous predictors (including both independent variables and moderators; X and Z in this case) will be mean-centered – i.e. the mean of the variable will be subtracted from it, so the new version has a mean of zero. Categorical moderator variables are discussed separately later.

Whatever the decision about centering, it is crucially important that the interaction term is calculated from the same form of the main variables that are entered into the regression; so if X and Z are centered to form new variables X_c and Z_c, then the term XZ is calculated by multiplying X_c and Z_c together (this interaction term is then left as it is, rather than itself being centered). The dependent variable, Y, is left in its raw form. An ordinary regression analysis can then be used with Y as the dependent variable, and X_c, Z_c and XZ as the independent variables.

For an example, let us consider a data set of 200 employees within a manufacturing company, with job performance (externally rated) as the dependent variable. We hypothesize that the relationship between training provided and performance will be stronger amongst employees whose roles have greater autonomy. Using *PERFORM*, *TRAIN* and *AUTON* to denote as the variables job performance, training provision and autonomy respectively, we would first create centered versions of *TRAIN* and *AUTON*, then create an interaction term (which I denote by *TRAXAUT*) by multiplying together the two centered variables, and then run the regression. Syntax for doing this in SPSS is in the appendix.

In this example, the (unstandardized) regression coefficients found may be as follows:

$$PERFORM = 3.20 + 0.62*TRAIN_c + 0.21*AUTON_c + 0.25*TRAXAUT$$

In order to determine whether the moderation is significant, we simply test the coefficient of the interaction term, 0.25; most software will do this automatically.[2] In our example, the coefficient has a standard error of 0.11, and therefore $p = 0.02$; the moderation effect is significant.

Some authors recommend that the interaction term is entered into the regression in a separate step. This is not necessary for the purposes of testing the interaction; however, it allows the computation of the increment in R^2 due to the interaction term alone. However for the remainder of the article I will assume all terms are entered together unless otherwise stated, and will comment more on hierarchical entry as part of the frequently asked questions at the end of the article.

Interpreting two-way interaction effects

The previous example gave us the result that the association between training and performance differs according to the level of autonomy. However, it is not entirely clear *how* it differs. The positive coefficient of the interaction term suggests that it becomes more positive as autonomy increases; however, the size and precise nature of this effect is not easy to divine from examination of the coefficients alone, and becomes even more so when one or more of the coefficients are negative, or the standard deviations of X and Z are very different.

To overcome this and enable easier interpretation, we usually plot the effect so we can interpret it visually. This is usually done by calculating predicted values of Y under different conditions (high and low values of the X, and high and low values of Z) and showing the predicted relationship ("simple slopes") between the X and Y at these different levels of Z.

[2] Technically, the test is to compare the ratio of the coefficient to its standard error with a *t*-distribution with 196 degrees of freedom: 196 because it is 200 (the sample size) minus the number of parameters being estimated (four: three coefficients for three independent variables, and one intercept).

Sometimes we might use important values of the variables to represent high and low values; if there is no good reason for choosing such values, a common method is to use values that are one standard deviation above and below the mean.

In our example above, the standard deviation of $TRAIN_c$ is 1.2, and the standard deviation of $AUTON_c$ is 0.9. The mean of both variables is 0, as they have been centered. Therefore the predicted value of *PERFORM* at low values of both $TRAIN_c$ and $AUTON_c$ is given by inserting these values into the regression equation shown earlier:

$$3.2 + 0.62*(-1.2) + 0.21*(-0.9) + 0.25*(-1.2*-0.9) = 2.54$$

Predicted values of *PERFORM* at other combinations of $TRAIN_c$ and $AUTON_c$ can be calculated similarly. However, various online resources can be used to do these calculations and plot the effects simultaneously: for example, www.jeremydawson.com/slopes.htm.

Figure 1 shows the plot of this effect created using a template at this web site. It demonstrates that the relationship between training and performance is always positive, but it is far more so for employees with greater autonomy (the dotted line) than for those with low autonomy (the solid line). Some researchers choose to use three lines, the additional line indicating the effect at average values of the moderator. This is equally correct and may be found preferable by some; however in this article I use only two lines for the sake of consistency (when advancing to three-way interactions, three lines would become unwieldy).

This may be sufficient for our needs; we know that the slopes of the lines are significantly different from each other (this is what is meant by the significant interaction term; Aiken & West, 1991), and we know the direction of the relationship. However, we may want to know more about the specific relationship between training and performance at particular levels of autonomy, so for example we could ask the question: for employees with low autonomy, is

there evidence that training would be beneficial to their performance? This can be done by using simple slope tests (Cohen, Cohen, West and Aiken, 2003).

Simple slope tests are used to evaluate whether the relationship (slope) between X and Y is significant at a particular value of Z. To perform a simple slope test, the slope itself can be calculated by substituting the value of Z into the regression equation, i.e. the slope is $b_1 + b_3Z$, and the standard error of this slope is calculated by

$$SE_S = \sqrt{s_{11} + Z^2 s_{33} + 2Z s_{13}} \quad [3]$$

where s_{11} and s_{33} are the variances of the coefficients b_1 and b_3 respectively, and s_{13} is the covariance of the two coefficients. Most software does not give these variances and covariances automatically, but will have an option to include it in the output of a regression analysis.[3] The significance of a simple slope is then tested by comparing the ratio of the slope to its standard error, i.e.

$$\frac{b_1 + b_3 Z}{\sqrt{s_{11} + Z^2 s_{33} + 2Z s_{13}}} \quad [4]$$

with a *t*-distribution with $n - k - 1$ degrees of freedom, where k is the number of predictors in the model (which is three if no control variables are included).

An alternative, indirect method for evaluating simple slopes uses a transformation of the moderating variable (what Aiken & West, 1991, call the "computer" method). This relies on the interpretation of the coefficient b_1 in equation [2]: it is the relationship between X and Y when Z = 0; i.e. if this coefficient is significant then the simple slope when Z has the value 0 is significant. Therefore if we were to transform Z such that the point where we wish to evaluate the simple slope has the value 0, and recalculate the interaction term using this transformed

[3] Note that the variance of a coefficient can be taken from the diagonal of the coefficient covariance matrix, i.e. the variance of a coefficient with itself; alternatively, it can be calculated by squaring the standard error of that coefficient.

value, then a new regression analysis using these terms would yield a coefficient b_1 which represents that particular simple slope and tests its significance. In practice, if the value of Z where we want to evaluate the simple slope is Z_{ss}, we would create the transformed variable $Z_t = Z - Z_{ss}$, create the new interaction term XZ_t, and enter the terms X, Z_t and XZ_t into a new regression. This is often more work than the method previously described; however, this transformation method becomes invaluable for probing more complex interactions (described later in this article).

Again, online resources exist which make it easy to test simple slopes, including at www.jeremydawson.com/slopes.htm, and via specialist R packages such as "pequod" (Mirisola & Seta, 2013). However, this sometimes results in researchers using simple slope tests without much thought, and using arbitrary values of the moderator such as one standard deviation above the mean (the "pick-a-point" approach; Rogosa, 1980). Going back to the earlier example, it may be tempting to evaluate the significance of the simple slopes plotted, which are one standard deviation above and below the mean value of autonomy (3.7 and 1.9 respectively). If we do this for a low value of autonomy (1.9), we find that the simple slope is not significantly different from zero (gradient = 0.40, $p = .06$). However, if we were to choose the value 2.0 instead of 1.9 – which is probably easier to interpret – we would find a significant slope (gradient = 0.42, $p = .04$). Therefore the findings depend on an often arbitrary choice of values, and in general it would be better to choose meaningful values of the moderator at which to evaluate these slopes. If there are no meaningful values to choose, then it may be better that such a simple slope test is not conducted, as it is probably unnecessary. I expand on this point in the frequently asked questions later in this article.

One approach that is sometimes used to circumvent this arbitrariness is Johnson-Neyman (J-N) technique (Bauer & Curran, 2005). This approach includes different methods for describing the variability, or uncertainty, about the estimates produced by the regression, rather than simply using hypothesis testing to examine whether the effect is different from zero or not. This includes the construction of confidence bands around simple slopes: a direct extension of the use of confidence intervals around parameters such as correlations or regression coefficients. More popular, though, is the evaluation of *regions of significance* (Aiken & West, 1991). These seek to identify the values of Z for which the X-Y relationship would be statistically significant. This can be helpful in understanding the relationship between X and Y, insomuch as it indicates values of the moderator at which the independent variable is more likely to be important; however, it still needs to be interpreted with caution. In our example, we would find that the relationship between training and job performance would be significant for any values of autonomy higher than 1.94. There is nothing special about this value of 1.94; it is merely the value, with this particular data set, above which the relationship would be found to be significant. If the sample size were 100 rather than 200 the value would be 2.58, and if the sample size were 300, the value would be 1.72. Thus the size of the region of significance is dependent on the sample size (just as larger sample sizes are more likely to generate statistically significant results), and the boundaries of the region do not represent estimates of any meaningful population parameters. Nevertheless, if interpreted correctly the region of significance is of greater use than simple slope tests alone. I do not reproduce the tests behind the regions of significance approach here, but an excellent online resource for this testing is available at http://quantpsy.org/interact/index.html.

Multiple moderators

Of course, even a situation of an X-Y relationship moderated by a single variable is somewhat simplistic, and very often life is more complicated than that. In our example, the relationship between training provided and job performance is moderated by autonomy, but it may also be moderated by experience: e.g. younger workers may have more to learn from training. Moreover, the moderating effects of experience may itself depend on autonomy; even if the worker is new, they may not be able to transfer the training to their job performance if they have little autonomy, but if they have a very autonomous role and are inexperienced, the effects of training may be exacerbated beyond that predicted by either moderator alone.

Statistically, this is represented by the following extension to equation [2] from earlier in this article:

$$Y = b_0 + b_1X + b_2Z + b_3W + b_4XZ + b_5XW + b_6WZ + b_7XZW + \varepsilon \quad [5]$$

where W is a second moderator (experience in our example). Importantly, this involves the main effects of each of the three predictor variables (the independent variable and the two moderators) and the three two-way interaction terms between each pair of variables as well as the three-way interaction term. The inclusion of the two-way interactions is crucial, as without these (or the main effects) the results could not be interpreted meaningfully. As before, the variables can be in their raw form, or centered, or *z*-standardized.

The significance of the three-way interaction term (i.e. the coefficient b_7) determines whether the moderating effect of one variable (Z) on the X-Y relationship is itself moderated by (i.e. dependent on) the other moderator, W. If it is significant then the next challenge is to interpret the interaction. As with the two-way case, a useful starting point is to plot the effect. Figure 2 shows a plot of the situation described above, with the relationship between training

provision and job performance moderated by both autonomy and experience.[4] This plot reveals a number of the effects mentioned. For example, the effect of training on performance for low-autonomy workers is modest whether the individuals are high in experience (slope 3; white squares) or low in experience (slope 4; black squares). The effect is somewhat larger for high autonomy workers who are also high in experience, but the effect is greatest for those who are low in experience and have highly autonomous jobs.

Following the plotting of this interaction, however, other questions may arise. Is training still beneficial for low autonomy workers? For more experienced workers specifically, does the level of autonomy affect the relationship between training provision and job performance? These questions can be answered either using simple slope tests (in the case of the former example), or slope difference tests (in the case of the latter).

Simple slope tests for three-way interactions are very similar to those for two-way interactions, but with more complex formulas for the test statistics. At given values of the moderators Z and W, the test of whether the relationship between X and Y is significant uses the test statistic:

$$\frac{b_1+b_4Z+b_5W+b_7ZW}{\sqrt{s_{11}+Z^2s_{44}+W^2s_{55}+Z^2W^2s_{77}+2(Zs_{14}+Ws_{15}+ZWs_{17}+ZWs_{45}+Z^2Ws_{47}+ZW^2s_{57})}} \quad [6]$$

where b_1, b_4 etc. are the coefficients from equation [6], and s_{11}, s_{14} etc. are the variances and covariances respectively of those coefficients. This test statistic is compared with a *t*-distribution with $n - k - 1$ degrees of freedom, where k is the number of predictors in the model (which is seven if no control variables are present).

[4] Template for plotting such effects, along with the simple slope and slope difference tests described later, are available at www.jeremydawson.com/slopes.htm.

This can be evaluated at any particular values of Z and W; as with two-way interactions, though, the significance will depend on the precise choice of these values, and so unless there is a specific theoretical reason for choosing values the helpfulness of the test is limited. For example, you may choose to examine whether training is beneficial for workers with ten years' experience and an autonomy value of 2.5; however, this does not tell you what the situation would be for workers with, say, 11 years' experience and autonomy of 2.6, and so unless the former situation is especially meaningful, it is probably too specific to be particularly informative. Likewise regions of significance can be calculated for these slopes (Preacher et al., 2006), but the same caveats as described for the two-way case apply here also.

More useful sometimes is a slope difference test: a test of whether the difference between a pair of slopes is significant (Dawson & Richter, 2006). For example, given an employee with ten years' experience, does the level of autonomy make a difference to the association between training and job performance? (This is the difference between slopes 1 and 3 in figure 2.) In the case of two-way interactions there is only one pair of slopes; the significance of the interaction term in the regression equation indicates whether or not these are significantly different from each other. For a three-way interaction, though, there are six possible pairs of lines, and there are six different formulas for the relevant test statistics, shown in table 1. Four of these rely on specific values of one or other moderator (e.g. ten years' experience in our example), and as such the result may also vary depending on the value chosen – although as they are not dependent on the value of the other moderator this is a lesser problem than with simple slope tests. The other two – situations where both moderators vary – do not depend on the values of the moderators at all, and as such are "purer" tests of the nature of the interaction. For example, is training more

effective for employees of high autonomy and high experience, or low autonomy and low experience? Such questions may not always be relevant for a particular theory however.

Sometimes it might be possible to find a three-way interaction significant but no significant slope differences. This is likely to be because it is not the X-Y relationship that is being moderated, but the Z-Y or W-Y relationship instead. Mathematically, it is equivalent whether Z and W are the moderators, X and W are the moderators, or X and Z are the moderators. So it may be beneficial in these cases to swap the predictor variables around; for example, it could be that the relationship between autonomy and job performance is moderated by training provided and experience. This should only be done if the alternative being tested makes sense theoretically; sometimes independent variables and moderators are almost interchangeable according to theory, but other times it is absolutely clear which should be which.

With all of these situations, I have described how the interaction can be "probed" after it has already been found significant. This post-hoc probing is, in many ways, atheoretical; if a hypothesis of a three-way interaction has been formed, it should be possible to specify exactly what form the interaction should take, and therefore what pairs of slopes should be different from each other. I strongly recommend that any three-way interaction that is hypothesized is accompanied by a full explanation of how the interaction should manifest itself, so it is known in advance what tests will need to be done. Examples of this are given in Dawson and Richter (2006).

It is also possible to extend the same logic and testing procedures to higher-order interactions still. Four-way interactions are occasionally found in experimental or quasi-experimental research with categorical predictors (factors), but are very rare indeed with continuous predictors. This is partly due to the methodological constraints that testing such

interactions would bring (larger sample sizes, higher reliability of original variables necessary), and partly due to the difficulty of theorizing about effects that are so conditional. However, the method of testing for such interactions is a direct extension of that for three-way interactions, and the examination of simple slopes and slope differences can be derived in the same way – albeit with considerably more complex equations.

It is also possible to have multiple moderators that do not interact with each other – that is, there are separate two-way interactions but no higher-order interactions. I will deal with such situations later, as one of the twelve frequently asked questions about moderation.

Testing and interpreting moderation in other types of regression models

Curvilinear effects

So far I have only described interactions that follow a straightforward linear regression relationship: the dependent variable (Y) is continuous, and the relationship between it and the independent variable (X) is linear at all values of the moderator(s). However, sometimes we need to examine different types of effects. We first consider the situation where Y is continuous, but the X-Y relationship is non-linear.

There are different ways in which non-linear effects can be modeled. In the natural sciences patterns may be found which are (for example) logarithmic, exponential, trigonometric or reciprocal. Although it is not impossible for such relationships to be found in management (or the social and behavioral sciences more generally), the relatively blunt measurement instruments used mean that a quadratic effect usually suffices to model the relationship; this can account for a variety of different types of effect, including U-shaped (or inverted U-shaped) relationships, or

those where the effect of X on Y increases (or decreases) more at higher, or lower, values of X. Therefore in this article I limit the discussion of curvilinear effects to quadratic relationships.

A straightforward extension of the regression equation [1] to include a quadratic element is

$$Y = b_0 + b_1X + b_2X^2 + \varepsilon \quad [7]$$

where X^2 is simply the independent variable, X, squared. With different values of b_0, b_1 and b_2 this can take on many different forms, describing many types of relationships; some of these are exemplified in figure 3. Of course, though, the form (and strength) of the relationship between X and Y may depend on one or more moderators. For a single moderator, Z, the regression equation becomes

$$Y = b_0 + b_1X + b_2X^2 + b_3Z + b_4XZ + b_5X^2Z + \varepsilon \quad [8]$$

Testing whether or not Z moderates the relationship between X and Y here is slightly more complicated. Examining the significance of the coefficient b_5 will tell us whether the curvilinear portion of the X-Y relationship is altered by the value of Z – in other words, whether the *form* of the relationship is altered. However, that does not answer the question of whether the *strength* of the relationship between X and Y is changed by Z; to do this we need to jointly test the coefficients b_4 and b_5. This can be done using an F-test between regression models – i.e. the complete model, and one without the XZ and X^2Z terms included – and can be accomplished easily in most standard statistical software. As with previous models, example syntax can be found in the appendix.

Returning to our training and job performance example, we might find that training has a strong effect on job performance at low levels, but above a certain level has little additional

benefit. This is easily represented with a quadratic effect. When autonomy is included as a moderator of this relationship, the equation found is:

$$PERFORM = 3.65 + 0.40*TRAIN_c - 0.15*TRAINSQ + 0.50*AUTON_c + 0.25*TRAXAUT - 0.14*TRASXAUT$$

where *TRAINSQ* is square of $TRAIN_c$, *TRASXAUT* is *TRAINSQ* multiplied by $AUTON_c$, and as before, *TRAXAUT* is $TRAIN_c$ multiplied by $AUTON_c$. As usual, the best way to begin interpreting the interaction is to plot it; this is shown in figure 4.

In common with linear interactions, much can be taken from the visual study of this plot. For example we can see that for individuals with high autonomy, there is a sharp advantage in moving from low to medium levels of training, but a small advantage at best in moving from moderate to higher levels of training. However, for low autonomy workers there is a far more modest association between training and job performance which is almost linear in nature. Thus the nature of the training-performance relationship (not just the strength) changes depending on the level of autonomy.

An obvious question, analogous to simple slopes in linear interactions, might be: is the X-Y relationship significant at a particular value of Z? This is a more problematic question than a simple slope analysis, however, because it is important to distinguish between the question of whether there is a significant *curvilinear* relationship at that value of Z, and the question of whether there is any relationship at all. The relevant test depends on this distinction. Either way we start by rearranging equation [8] above:

$$Y = b_0 + b_3Z + (b_1 + b_4Z)X + (b_2 + b_5Z)X^2 \quad [9]$$

If there were a quadratic relationship, the value of ($b_2 + b_5Z$) would be non-zero. Therefore a test of a curvilinear relationship would examine the significance of this term: using the same logic as tests described earlier, this means comparing the test statistic

$$\frac{b_2+b_5Z}{\sqrt{s_{22}+Z^2s_{55}+2Zs_{25}}} \quad [10]$$

with a *t*-distribution with $n - k - 1$ degrees of freedom, where k is the number of predictors in the model (including the five in equation 5 and any control variables). If this term is significant then there is a curvilinear relationship between X and Y at this value of Z. A more relevant test, however, may be of whether there is any relationship between X and Y (linear or curvilinear) at this value of Z. In this situation we can revert to the method described earlier: centering the moderator at the value at which we wish to test the X-Y relationship. If this is done, then the b_1 and b_2 terms between them describe whether there is any relationship (linear or curvilinear) between X and Y at this point. Therefore, though somewhat counterintuitive, the best way to test this would be to examine whether the X and X^2 terms add significant variance to the model after the inclusion of the other terms (Z, XZ and X^2Z). This will only test the effect for the value of the moderator around which Z is centered. However, unless there is a specific hypothesis regarding the curvilinear nature of the association between X and Y at a particular value of Z, this latter, more general, form of the test is recommended.

Sometimes you might expect a linear relationship between X and Y, but a curvilinear effect of the moderator. The appropriate regression equation for this would be

$$Y = b_0 + b_1X + b_2Z + b_3Z^2 + b_4XZ + b_5XZ^2 + \varepsilon \quad [11]$$

and this can be tested in exactly the same way as before, with the significance of b_5 indicating whether there is curvilinear moderation. The interpretation of such results is more difficult, however, because a plot of the relationship between X and Y will show only linear relationships;

showing high and low values of the moderator will give no indication of the lack of linearity in the effect. Therefore, in such situations it is better to plot at least three lines, as in figure 5. This reveals that where autonomy is low, there is little relationship between training provision and job performance; a moderate level of autonomy actually gives a negative relationship between training and performance, but at a high level of autonomy there is a strong positive relationship between training and performance. If only two lines had been plotted, the curvilinear aspect of this changing relationship would have been missed entirely. The test of a simple slope for this situation has the test statistic

$$\frac{b_1+b_4Z+b_5Z^2}{\sqrt{s_{11}+Z^2s_{44}+Z^4s_{55}+2(Zs_{14}+Z^2s_{15}+Z^3s_{45})}} \quad [12]$$

and this is again compared with a *t*-distribution with $n - k - 1$ degrees of freedom.

Sometimes researchers specifically hypothesize curvilinear effects – either of the independent variable or the moderator. However, even when this is not hypothesized it can be worth checking whether such an effect exists, as linearity of the model is one of the assumptions for regression analysis. If a curvilinear effect exists but has not been tested in the model, this would usually manifest itself by a skewed histogram of residuals, and evidence of a non-linear effect in the plot of residuals against predicted values (see Cohen et al., 2003, for more on these plots). If such effects exist, then it is worth checking for curvilinear effects of both the independent variable and the moderator.

Of course, it is also possible to have curvilinear three-way interactions. The logic here is simply extended from the two separate interaction types – curvilinear interactions and three-way linear interactions – and the equation takes the form

$$Y = b_0 + b_1X + b_2X^2 + b_3Z + b_4W + b_5XZ + b_6X^2Z + b_7XW + b_8X^2W + b_9ZW + b_{10}XZW + b_{11}X^2ZW + \varepsilon \quad [13]$$

It is the term b_{11} that determines whether or not there is a significant curvilinear three-way interaction between the variables. As in the two-way case, simple curves can be tested for particular values of Z and W by centering the variables around these values, and testing the variance explained by the X and X^2 terms after the others have been included in the regression. The three-way case, though, brings up the possibility of testing for "curve differences" (analogous to slope differences in the linear case). For example, figure 6 shows a curvilinear, three-way interaction between training provision, autonomy and experience on job performance (training provision as the curvilinear effect). The plot suggests that training is most beneficial for high autonomy, low experience workers; however, it might be hypothesized specifically that there is a difference between the effect of training between these and high autonomy, high experience workers (i.e. for high autonomy employees, training has a differential effect between those with more and less experience). A full curve difference test has not been developed, as it would be far more complex than in the linear case. However, we can use the fact that one variable (autonomy) remains constant between these two curves. If we call autonomy Z and experience W in equation 7 above, and if we center autonomy around the high level used for this test (e.g. mean + one standard deviation), then the differential effect of experience on the training-performance relationship at this level of autonomy is given by the XW and X^2W terms. Therefore the curve difference test is given by the test of whether these terms add significant variance to the model after all other terms have been included (but only for the value of $Z = 0$). Again, syntax for performing this test in SPSS is given in the appendix.

Extensions to non-Normal outcomes

Just as the assumptions about the distribution of the dependent variable in OLS regression can be relaxed with generalized linear modeling (e.g. binary logistic regression, Poisson regression, probit regression), the same is true when testing moderation. A binary logistic moderated regression equation with a logit link function, for example, would be given by

$$\text{logit}(Y) = \log[Y/(1-Y)] = b_0 + b_1X + b_2Z + b_3XZ + \varepsilon \quad [14]$$

or, equivalently,

$$Y = \frac{e^{b_0+b_1X+b_2Z+b_3XZ+\varepsilon}}{1+e^{b_0+b_1X+b_2Z+b_3XZ+\varepsilon}} \quad [15]$$

where Y is the probability of a "successful" outcome. Given familiarity with the generalized linear model in question, therefore, testing for moderation is straightforward: it simply involves entering the same predictor terms into the model as with the "Normal" version. The significance of the interaction term determines whether or not there is moderation.

Interpreting a significant interaction, however, is slightly less straightforward. As with previously described cases, the best starting point is to plot the relationship between X and Y at low and high values of Z. However, because of the logit link function this cannot be done by drawing a straight line between two points; instead, an entire curve should be plotted for each chosen value of Z (even though the equation for Y involves only X and Z, the relationship between X and Y will usually appear curved). An approximation to the curve may be generated by choosing regular intervals between low and high values of X, evaluating Y from equation [15] at that value of X (and Z), and plotting the lines between each pair of values.[5]

For example, we might be interested in predicting whether employees would be absent due to illness for more than five days over the course of a year. We may know that employees

[5] This is the method used by the relevant template at www.jeremydawson.com/slopes.htm, where there are also appropriate templates for three-way interactions, and two- and three-way interactions with Poisson regression

with more work pressure are more likely to be absent; we might suspect that this relationship is stronger for younger workers. The logistic regression gives us the following estimates:

$$\text{logit}(ABSENCE) = -2.11 + 0.42*WKPRES_c - 0.29*AGE_c + 0.24*WKPXAGE$$

where $WKPRES_c$ is the centered version of work pressure, AGE_c is the centered version of age, and *WKPXAGE* is the interaction term between these two. *ABSENCE* is a binary variable with the value 1 if the individual was absent from work for more than five days in the course of the year following the survey, and 0 otherwise.

A plot of the interaction is shown in figure 7. As with OLS regression, this plots the expected values of Y for different values of X, and at high and low values of Z. Here, the expected values of Y means the probability that an individual is absent for more than five days; the expected value is calculated by substituting the relevant values of work pressure and age into the formula

$$E(\text{Y}) = \frac{e^{-2.11 + 0.42*WKPRES_c - 0.29*AGE_c + 0.24*WKPRES_c*AGE_c}}{1 + e^{-2.11 + 0.42*WKPRES_c - 0.29*AGE_c + 0.24*WKPRES_c*AGE_c}}$$

In this plot, the values of X (work pressure) range from 1.5 standard deviations below the mean to 1.5 standard deviations above, with values plotted every 0.25 standard deviations and a line drawn between these to approximate the curve. It can be seen that the form of the interaction is not so straightforwardly determined from the coefficients as it is with OLS regression. However, the signs of the coefficients are still helpful; the positive coefficient for work pressure means that more work pressure is generally associated with more absence, the negative coefficient for age means that younger workers are more likely to be absent than older workers, and the positive interaction coefficient means that the effect is stronger for older workers. A plot, however, is essential to see the precise patterns and extent of the curvature.

Extensions to three-way interactions are possible using the same methods. In both the two-way and three-way cases, an obvious supplementary question to ask may be whether the relationship between X and Y is significant at a particular value of the moderator(s) – a direct analogy to the simple slope test in moderated OLS regression. As with curvilinear interactions with Normal outcomes, the best way of testing whether the simple slopes/curves (i.e. the relationship between X and Y at particular values of the moderator) are significant is to center the moderator around the value to be tested before commencing; then the significance of the b_1 term gives a test of whether the slope is significant at that value. Some slope difference tests for three-way interactions can also be run in this way. For those pairs of slopes where one moderator remains constant (cases a to d in table 1), that variable can be centered around that constant value, and then the slope difference test is given by the resulting significance of the interaction between the other two terms. For example, if we want to test for the difference of the slopes for high Z, high W and high Z, low W, then we center Z around its high value, calculate all the interaction terms and run the three-way interaction test, and the slope difference test is given by the significance of the XW term.

Extensions to other non-Normal outcomes are similar. For example, if the dependent variable is a count score, Poisson or negative binomial regression may be suitable. The link function for these is usually a straightforward log link (rather than the logit, or log-odds, link used for binary logistic regression) – which makes the interpretation slightly easier – but otherwise the method is directly equivalent. Say, for example, we are interested in counting the number of occasions on which an employee is absent, but otherwise use the same predictors as in the previous example. We find the following result:

$$\log(TIMESABS) = 0.23 + 0.82*WKPRES_c - 0.09*AGE_c + 0.20*WKPXAGE$$

In this case, the expected values of TIMESABS (the number of occasions an employee is absent) is given by the formula

$$e^{0.23 + 0.82*WKPRES_c - 0.09*AGE_c + 0.20*WKPRES_c*AGE_c}$$

because the exponential is the inverse function of the log link. This effect is plotted in figure 8. Extensions to three-way interactions, simple slope tests and slope difference tests follow in the same way as for binary logistic regression.

Note that the methods for plotting and interpreting the interaction depend only on the link function, not the precise distribution of the dependent variable, so exactly the same method could be used for either Poisson or negative binomial regression. Testing interactions using other generalized linear models – e.g. probit regression, ordinal or multinomial logistic regression – can be done in an equivalent way.

Twelve frequently asked questions concerning testing of interactions

1. Does it matter which variable is which?

It is clear from equation [2] that the independent variable and moderator can be swapped without making any difference to the regression model – and specifically to the test of whether an interaction exists. Therefore, mathematically, Z moderating the relationship between X and Y is identical to X moderating the relationship between Z and Y. The decision about which variable should be treated as the independent variable is therefore a theoretical one; sometimes this might be obvious, but at other times this might be open to some interpretation. For example, in the model used at the start of this article, the relationship between training and performance was moderated by autonomy; this makes sense if the starting point for research is understanding the effect of training on job performance, and determining when it makes a difference. However,

another researcher might have job design as the main focus of their study, hypothesizing that employees with more autonomy would be better able to perform well, and that this effect is exacerbated when they have had more training. In both cases the data would lead to the same main conclusion – that there is an interaction between the two variables – but the way this is displayed and interpreted would be different.

Sometimes this symmetry between the two variables can be useful. When interpreting this interaction (shown in figure 1) earlier in the article, we focused on the relationship between training and performance: this was positive, and became more positive when autonomy was greater. But a supplementary question that could be asked is about the difference in performance when training is high; in other words, are the two points on the right of the plot significantly different from each other? This is not a question that can be answered directly by probing this plot; however, be re-plotting the interaction with the independent variable and moderator swapped around, we can look at the difference between these points as a slope instead (see figure 9). The question of whether the two points in figure 1 are different from each other is equivalent to the question of whether the "high training" slope in figure 9 (the dotted line) is significant. Thus, using a simple slope test on this alternative plot is the best way of answering the question.

For three-way interactions, the same is true, but there are three possible variables for the independent variable. For curvilinear interactions, however, this would be more complex; if a curvilinear relationship between X and Y is moderated (linearly) by Z, then re-plotting the interaction with Z as the independent variable would result in straight lines, and thus the curvilinear nature of the interaction would be far harder to interpret.

While the independent variables and moderators are interchangeable, however, the same is definitely not the case with the dependent variable. Although in a simple relationship

involving X and Y there is some symmetry there, once the moderator Z is introduced this is lost; in other words, determining which of X, Y and Z is the criterion variable is of critical importance. For further discussion on this see Landis and Dunlap (2000).

2. Should I center my variables?

Different authors have made different recommendations regarding the centering of independent variables and moderators. Some have recommended mean-centering (i.e. subtracting the mean from the value of the original variable so that it has a mean of 0); others *z*-standardization (which does the same, and then divides by the standard deviation, so that it has a mean of 0 and a standard deviation of 1); others suggest leaving the variables in their raw form. In truth, with the exception of cases of extreme multicollinearity, the decision does not make any difference to the testing of the interaction term; the *p*-value for the interaction term and the subsequent interaction plot should be identical whichever way it is done (Dalal & Zickar, 2012; Kromrey & Foster-Johnson, 1998). Centering does, however, make a difference to the estimation and significance of the other terms in the model: something we use to our advantage in the indirect form of the simple slope test, as the interpretation of the X coefficient (b_1 in equation [2]) is the relationship between X and Y when Z = 0. Therefore, unless the value 0 is intrinsically meaningful for an independent variable or moderator (e.g. in the case of a binary variable), I recommend that these variables are either mean-centered or *z*-standardized before the computation of the interaction term.

The choice between mean centering and *z*-standardization is more a matter of personal preference. Both methods will produce identical findings, and there are some minor advantages to each. Mean centering the variables will ensure that the (unstandardized) regression

coefficients of the main effects can be interpreted directly in terms of the original variables. For many people this is reason enough to use this method. On the other hand, *z*-standardization allows easy interpretation of the form of the interaction by addition and subtraction of the coefficients, makes formulas for some probing methods more straightforward, and is easily accomplished with a simple command in SPSS (Dawson & Richter, 2006).

Whichever method is chosen, there are some rules that should be obeyed. First, it is essential to create the interaction term using the same versions of the independent variable and moderator(s) that are used in the analysis. The interaction term itself should not be centered or *z*-standardized (this also applies to the two-way interaction terms when testing three-way interactions, etc.). The regression coefficients that are interpreted are always the unstandardized versions. Also, it is highly advisable to mean center or (*z*-standardize) any independent variables in the model that are not part of the interaction being tested (e.g. control variables), as otherwise the predicted values are more difficult to calculate (and the plots produced by some automatic templates will display incorrectly). Finally, the dependent variable (criterion) should *not* be centered or *z*-standardized; doing so would result in interpretations and plots that would fail to reflect the true variation in that variable.

3. What if my moderator is categorical?

Sometimes our moderator variable may be categorical in nature, for example gender or job role. If it has only two categories (a binary variable) then the process of testing for moderation is almost identical to that described earlier; the only difference is that it would not be appropriate to center the moderator variable (however, coding the values as 0 and 1 makes the interpretation somewhat easier). It would also be straightforward to choose the values of the

moderator at which to plot and test simple slopes: these would necessarily be the two values that the variable can take on.[6] However, if a moderator has more than two categories, testing for moderation is more complex. This can be done either using an ANCOVA approach (see e.g. Rutherford, 2001), or within regression analysis using dummy variables (binary variables indicating whether or not a case is a member of a particular category).

In this latter approach, dummy variables are created for all but one of the categories of the moderator variable (the one without a dummy variable is known as the reference category). These would be used directly in place of the moderator variable itself – so that if a moderator variable has k categories, then there would be $k - 1$ dummy variables entered into the regression as raw variables, and $k - 1$ separate interaction terms between these dummy variables and the independent variable. Some software (e.g. R) will automatically create dummy variables as part of the regression procedure if a categorical moderator is included; however, much other software (e.g. SPSS) does not, and so the dummy variables need to be created separately. Example syntax for creating dummy variables and testing such interactions is given in the appendix.

Plotting these interactions is slightly more difficult than plotting a straightforward two-way interaction because it requires more lines; specifically, there will be a separate line for each category of the moderator. The question of whether a line for a given category differs significantly from the line of the reference category is given by the significance of the interaction term for that particular category. If it is necessary to test for differences between two other lines, neither of which is for the reference category, then the regression would need to be re-run with one of these categories as the reference category instead.

[6] There is a specific template for binary moderators at www.jeremydawson.com/slopes.htm, as well as a generic template which allows any combination of binary and continuous independent and moderating variables.

An important consideration about categorical moderators is that they should only be used when the variable was originally measured as categories. Continuous variables should never be converted to categorical variables for the purpose of testing interactions. Doing so reduces the statistical power of the test, making it more difficult to detect significant effects (Stone-Romero & Anderson, 1994; Cohen, Cohen, West & Aiken, 2003), as well as throwing up theoretical questions about why particular dividing points should be used.

4. Should I use hierarchical regression?

As mentioned earlier in the article, sometimes researchers may choose to enter the variables in a hierarchical manner, i.e. in different steps, with the interaction term being entered last. This may be as much to do with tradition as anything else, because there is limited statistical rationale for doing it this way. Certainly, if the interaction term is significant, then it does not make sense to interpret versions of the model that do not include it, as those models will be mis-specified and therefore violating an assumption of regression analysis. (Of course if the interaction term is not significant, then removing it and interpreting the rest of the model may be a sensible approach.)

Why then might you choose to use hierarchical entry of predictor variables? There are at least two reasons why this might be appropriate. First, in order to calculate the effect size for the interaction (see question 5), it is necessary to know the increment in R^2 due to the interaction. This could be achieved either by entering the interaction term after the rest of the model, or by removing the interaction term from the full model in a second step. Second, in order to perform the version of the simple slope test I recommended for moderation of curvilinear effects, it is

necessary to enter the X and X^2 terms in a final step after the Z, XZ and X^2Z terms. However, only the full model in each case would be plotted and interpreted.

5. Should I conduct simple slope tests, and if so, how?

In some sections of the literature, simple slope tests have almost become "de rigueur" – a significant interaction is seldom reported without such tests. Sometimes these are requested by reviewers even when the author chooses not to include them. This is unfortunate, because simple slope tests in themselves often tell us little about the effect being studied.

A simple slope test is a conditional test; in the simple two way case, this tests whether there is a significant association between X and Y at a particular value of Z. The fact that this is about a particular value of Z is paramount to the interpretation of the test, although this is often overlooked. For example, a common supplement to the testing of the interaction effect plotted in figure 1 would be performing simple slope tests on the two lines shown in the plot. However, this merely tells us whether there is evidence of a relationship between training and job performance when autonomy has the value 1.9 or 3.7. There is nothing particularly meaningful about these values; they are purely arbitrary examples of relatively low and high levels of autonomy, and neither has any real intrinsic meaning nor represents generic low and high levels. Therefore the significance or otherwise of these slopes is indicative of something of very limited use.

Moreover, such testing – particularly for two-way interactions – often overshadows the significance of the interaction itself; if the interaction term is significant, then that immediately tells us that the association between X and Y differs significantly at different levels of Z. Often

this is sufficient information (along with the effect size/plot) to tell us what we need to know about whether the hypothesis is supported.

This is not to say that simple slope tests should never be used. However, only in certain circumstances would it be necessary to test the association between X and Y at a particular value of Z. Such situations are clear in the case of categorical moderators: for example, testing whether training is related to job performance for different job groups. Even for continuous moderators, however, this might be beneficial: for example, if it is known that the autonomy level 4.0 corresponds to the level where employees can make most key decisions about their own work, then testing the association at this level would be meaningful. Likewise, if age is the moderator, then testing the association when age = 20 or age = 50 would give meaningful interpretations of the relationship for people of these ages. However, such insight is rarely gained by automatically choosing values one standard deviation above and below the mean.

If you do conduct simple slope tests, then there are two main methods for doing this. Detail of these for the different models is given in the relevant earlier section of this article; broadly, though, the direct method is to test the significance of a specific combination of regression coefficients and coefficient covariances; the indirect method involves centering the moderator around the value to be tested, and re-running the regression using this new version of the variable. The indirect method is somewhat more long-winded, but has the advantage of being applicable in non-linear forms of moderation as well, where no precise equivalent of the direct method exists.

6. How do I measure the size of a moderation effect?

It is generally recognized that R^2 is not an ideal metric for measuring the size of an interaction effect, due to the inevitability of shared variance between the X, Z and XZ terms. Rather, it is more helpful to examine f^2, the ratio of variance explained by the interaction term alone to the unexplained variance in the final model:

$$f^2 = \frac{{R^2}_2 - {R^2}_1}{1 - {R^2}_2} \quad [16]$$

where R^2_1 and R^2_2 represent the variance explained by the models including and excluding the interaction term respectively (Aiken & West, 1991). Aguinis, Beaty, Boik and Pierce (2005) found that, for categorical variables, the values of f^2 found in published research were very low: a median of .002 across 30 years' worth of articles in three leading journals in management and applied psychology. For comparison, Cohen et al. (2003) describe .02 as being a small effect size. It is clear, therefore, that many studies which find significant interaction effects only have small effect sizes. As researchers it is important to acknowledge this, and focus on the practical relevance of findings rather than their statistical significance alone, e.g. by determining whether the extent of change in the association between X and Y due to a (say) one standard deviation change in Z is something that would be clearly relevant to the participants. It may also help to use relative importance analysis (Tonidandel & LeBreton, 2011) to demonstrate the relative importance of the interaction effect compared with the main effects.

7. What sample size do I need?

Choosing a sample size for any analysis is far from an easy process; although rules of thumb are sometimes used for certain types of analysis, these are generally approximations based on assumptions that may or may not be appropriate. All decisions are informed by statistical power (the ability to detect effects where they truly exist) and precision (the accuracy with which

parameters can be estimated; power and precision being two sides of the same coin). However, this becomes less straightforward the more complex the analysis is.

It is a well-known result that the power of testing interaction effects is generally lower than for testing main effects (McClelland & Judd, 1993). As with all types of analysis, sample size is the single biggest factor affecting power; however, a number of factors have been shown to reduce power for moderator effects specifically. For example, measurement error (lack of reliability) is a major source of loss of power; if both the independent variable and moderator suffer from this, then the measurement error in the interaction term is exacerbated (Dunlap & Kemery, 1988). Other factors known to attenuate power for detecting interactions include intercorrelations between the predictors (Aguinis, 1995), range restriction (Aguinis & Stone-Romero, 1997), scale coarseness and transformation of non-Normal outcomes (Aguinis, 1995, 2004), differing distributions of variables (Wilcox, 1998), and artificial categorization of continuous variables (Stone-Romero & Anderson, 1994). The presence of any of these issues will increase the sample size needed to achieve the same power.

So the answer to the question "*What sample size do I need?*" is difficult to give, other than to say it is probably substantially larger than for non-moderated relationships. However, Shieh (2009) gave some formulas to aid determination of sample sizes for specific levels of power and effect size. Although such sample size determination is rare in management research for interaction effects specifically, it is noteworthy that Shieh finds the sample size required to detect a relatively large effect (equivalent to an f^2 of around 0.3) with 90% power is about 137-154 cases (depending on the method of estimation used). By way of comparison, the sample size to detect a simple correlation with the same f^2 with the same level of power is just 41. Thus even

for simple two-way interactions without any significant attenuating effects, a considerable sample size is advisable.

8. Should I test for curvilinear effects instead of, or as well as, moderation?

As noted earlier in the article, curvilinear relationships between independent and dependent variables are not uncommon, and quadratic regression (regressing Y on X and X^2) is a relatively comprehensive tool to capture them. However, there is a danger that this could, instead, be picked up as an interaction between X and Z if the independent variable and moderator are correlated. As Cortina (1993) explains, the stronger the relationship between X and Z, the more likely a (true) curvilinear effect is to be (falsely) picked up as a moderated effect. Looking at it another way, a quadratic relationship in X is the same as a relationship between X and Y moderated by X – the nature of the association depending on the value of X. Despite this, testing for moderation is far more prevalent in management research than testing for curvilinear effects.

Therefore it is advisable that curvilinear effects be tested whenever there is a sizeable correlation between X and Z. Cortina (1993) suggests entering the terms X, Z, X^2 and Z^2 before entering the XZ term into the regression; as I pointed out above, the hierarchical entry is not strictly necessary here. Nevertheless, the inclusion of all five terms provides a conservative test of the interaction – if the XZ term is still significant despite the inclusion of the other terms, then there is likely to be a true moderating effect above and beyond any curvilinear effects. However, the conservative nature of the test means there is a relative lack of statistical power. As a result, I would suggest that such tests are advisable when there is a moderate correlation between X and Z (between 0.30 and 0.50), and essential when there is a large correlation (above 0.50).

The inclusion of all five terms means that this test is equivalent to second-order polynomial regression (Edwards, 2001; Edwards & Parry, 1993) – although this is known more for testing congruence between predictors, it is also appropriate for testing and interpreting interaction effects when the curvilinear effects are also found. If a significant interaction is found, but the X^2 and Z^2 terms are not significant, then it is more parsimonious to interpret the usual form of the interaction using the methods described earlier in this article. If both curvilinear and interaction terms are found to be significant, then it would often make sense to test for curvilinear moderation as described earlier in this article. For an introduction to polynomial regression in organizational research, see Shanock, Baran, Gentry, Pattison and Heggestad (2010).

9. What should I do if I have several independent variables all being moderated, or multiple moderators for a single independent variable?

Regression models including multiple independent variables (X_1, X_2 etc.) are common. If you introduce a moderator to these models, how should you test this? This scenario might involve, for example, the effect of personality on proactivity, moderated by work demands. Researchers commonly use the Big Five model of personality, and thus there would be five independent variables to consider here.

The situation is relatively straightforward to test, but more difficult to interpret, particularly if there are correlations between the independent variables. Ideally, the regression should include all independent variables, the moderator, and interactions between the moderator and each independent variable (a total of 11 variables in the scenario above). It is important in this situation that all predictors are mean-centered or *z*-standardized before the calculation of

interaction terms and the regression analysis. The initial test then depends on the precise hypothesis; for example, if the hypothesis were as general as "the relationship between personality and proactivity is moderated by work demands" then the required test would determine whether a significant increment in R^2 is made by the five interaction terms between them. More frequently, however, there may be separate hypotheses for different independent variables (e.g. "the relationship between neuroticism and proactivity is moderated by work demands"); this would allow individual coefficients of the relevant interactions to be tested. Non-significant interactions can also be removed from this model to allow optimal interpretation of the significant interactions; this helps to reduce multicollinearity. Each significant interaction can then be plotted and interpreted separately; importantly though, the interpretation of the interaction between X_1 and Z is at the mean level of all other independent variables X_2, X_3 etc.

The situation of multiple moderators is not dissimilar, except that there is a greater chance that the moderators would themselves interact, thus creating three-way or higher order interactions. For example, consider a relationship between X and Y which might be moderated by both Z_1 and Z_2. The test of this would be a regression analysis including the terms X, Z_1, Z_2, XZ_1 and XZ_2. It is possible that XZ_1 is significant and XZ_2 is not (so Z_1 moderates the relationship but Z_2 does not) or vice versa, or indeed that both are significant. However, if both Z_1 and Z_2 affect the relationship between X and Y then it would be natural also to test the three-way interaction between X, Z_1 and Z_2. If this three-way interaction is significant then it should be interpreted using the methods described earlier in this article. If it is not, then the interactions are most easily interpreted separately, again with the interpretation being at mean levels of the other (potential) moderator – the lack of a three-way interaction means that such an interpretation is reasonable.

10. Should I hypothesize the form of my interactions in advance?

In a word: yes! The methods described in this article have relied on testing the significance of interaction effects: i.e. null hypothesis significance testing (NHST). NHST has some detractors, and the use of methods such as effect size testing (EST) with confidence intervals to replace or supplement NHST is growing in management and psychology (Cortina & Landis, 2011); nevertheless, even methods such as EST and the Johnson-Neyman technique (Bauer & Curran, 2005) are based on an underlying expectation that an effect exists. Therefore not only should the existence of an interaction effect be predicted, but also its form. In particular, whether a moderator increases or decreases the association between two other variables should be specified as part of the a priori hypothesis. For two-way interactions this is relatively straightforward; simply stating the direction of the interaction is often sufficient, although sometimes suggesting whether the main X-Y effect would be positive, negative or null at high and low values of the moderator may be beneficial too. This would require stating what is meant by "high" and "low" values, and this would give rise to meaningful simple slope tests at these values.

For three-way (and higher) interactions, however, this requires more work. In particular, it is advisable to hypothesize how the lines in a plot of such an interaction should differ. This then enables the a priori specification of which slope difference tests should be used, and reduces the likelihood of a type I error (an incorrect significant effect). For examples of how this might be done, see Dawson and Richter (2006).

11. Do these methods work with multilevel models?

Within multilevel models, the method of testing the interactions themselves is directly equivalent to the methods explained above. This is regardless of whether the independent variables (including moderators) exist at level 1 (e.g. individuals), level 2 (e.g. teams), or a mixture between the two (cross-level interactions). The precise methods of testing such effects are covered in detail in many other texts (see for example Pinheiro & Bates, 2000; Snijders & Bosker, 1999; West, Welch & Galecki, 2006), and depend on the software used.

It is worth noting, however, that the interpretation of such interactions may be less straightforward. Effects can usually be plotted using the same templates as for single-level models, but further probing is more complicated. The formulas for simple slope tests, slope difference tests and regions of significance do not apply, and unless models contain no random effects, only approximations of these tests exist (Bauer & Curran, 2005). The indirect version of the simple slope test described earlier in this article can still be used, however. Tools for probing such interactions – both two-way and three-way – can be found at http://quantpsy.org/interact/index.html.

12. Can I test moderation within more complex types of model?

Yes, it is generally possible to test for moderators within more sophisticated modeling structures, although there is some variation in extent to which different types of models can currently incorporate different elements of moderation testing (e.g. regions of significance, non-Normal outcomes).

A good example of this is combining moderation with mediation. Two distinct (but conceptually similar) methods for testing such models were developed simultaneously by Preacher, Rucker and Hayes (2007), and by Edwards and Lambert (2007). Both methods

evaluate the conditional effect of X on Y via M at different levels of the moderator, Z. The detail of these methods is not reproduced here, but for further details see the original papers; online resources to help with the testing and interpretation of these effects can be found at http://quantpsy.org/medn.htm for Preacher et al.'s method, and at http://public.kenan-flagler.unc.edu/faculty/edwardsj/downloads.htm for Edwards and Lambert's method. A good summary of mediation in organizational research, including combining mediation and moderation, is given by MacKinnon, Coxe and Baraldi (2012).

Other recent developments have enabled the testing of latent interaction effects in structural equation modeling without having to create interactions between individual indicators of the variables (Klein & Moosbrugger, 2000; Muthén & Muthén, 1998-2011), which partially circumnavigates the problem of decreasing reliability of interaction terms (Jaccard & Wan, 1995). This is particularly relevant when the independent variable and/or moderator are formed of questionnaire scale items. Meanwhile, there is a large literature on the specific issues with categorical moderator variables; for example methods have been developed to control for heterogeneity of variance across groups (Aguinis et al., 2005; Overton, 2001). Likewise methods exist for testing interaction effects in multilevel and longitudinal research, and interactions in meta-analysis.

Conclusions

This article has described the purpose of, and procedure for, testing and interpreting interaction effects involving moderator variables. Such tests are already well-used within management and psychology research; however, often they may not be used to their full potential, or understood fully. The expansion to curvilinear effects and non-Normal outcome

variables should enable some researchers to test and interpret effects in a way that might not have been possible previously. The description of simple slope tests, slope difference tests and other probing techniques should clarify some matters about how and when these can be done, and the limitations of their use. The answers to twelve frequently asked questions (which reflect questions often posed to me over the last ten years) should help give some guidance on issues that researchers may be unclear about.

There is still much to be learned about moderation, however. Although the basic linear models for Normal outcomes are well-established, there is more to be learned about testing and probing non-linear relationships (particularly beyond the relatively simple quadratic effects described in this article), and for non-Normal outcomes and non-standard data structures. Possible directions for future research in this area include the development of probing techniques (e.g. to allow more accurate estimation of confidence intervals) with such models, further investigation into power and sample size calculations for non-standard models, and the development of effect size metrics for non-Normal outcomes.

References

Aguinis, H. (1995). Statistical power problems with moderated regression in management research. *Journal of Management, 21*, 1141–1158.

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: Guilford Press.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82,* 192–206.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90,* 94–107.

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

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

Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management, 19,* 915-922.

Cortina, J. M., & Landis, R. S. (2011). The earth is not round (p = .00). *Organizational Research Methods, 14,* 332-349.

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods, 15,* 339-362.

Dawson, J. F. & Richter, A. W. (2006). Probing three-way interactions: The development and application of a slope difference test. *Journal of Applied Psychology, 91,* 917-926.

Dunlap, W. P., & Kemery, E. R. (1988). Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes, 41,* 248-258.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265–287.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12,* 1-22.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348-357.

Johnson, R. E., Rosen, C. C., & Chang, C. (2011). To aggregate or not to aggregate: Steps for developing and validating higher-order multidimensional constructs. *Journal of Business and Psychology, 26,* 241–248.

Klein, A. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65,* 457-474.

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

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business and Psychology*, DOI 10.1007/s10869-013-9285-x.

Landis, R. S., & Dunlap, W. P. (2000). Moderated multiple regression tests are criterion specific. *Organizational Research Methods, 3,* 254-266.

Locke, E. A., Shaw, K. N., Saari, L. M.., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*, 125–152.

MacKinnon, D. P., Coxe, S., & Baraldi, A. N. (2012). Guidelines for the investigation of mediating variables in business research. *Journal of Business and Psychology, 27,* 1–14.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376.

Mirisola, A., & Seta, L. (2013). *pequod: Moderated regression package. R package version 0.0-3* [Computer software]. Retrieved March 20, 2013. Available from http://CRAN.R-project.org/package=pequod

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Overton, R. C. (2001).Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods, 6,* 218–233.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS. Statistics and computing*. New York: Springer.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational & Behavioral Statistics, 31,* 437-448.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42,* 185-227.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88,* 307–321.

Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM approach*. London: Sage.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25,* 543–554.

Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables: Power and sample size considerations. *Organizational Research Methods, 12,* 510-528.

Snijders, T., & Bosker, R. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd edition).* Thousand Oaks, CA: Sage.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology, 79,* 354–359.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology, 26,* 1–9.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89*, 1008-22.

West, B., Welch, K. B., & Galecki, A. T. (2007). Linear Mixed Models: A Practical Guide Using Statistical Software. Boca Raton, FL: Chapman & Hall.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53,* 300–314.

Table 1. Test statistics for differences between slopes in a three-way interaction.

	Slopes	Test statistic
(a)	(1) and (2)	$t = \dfrac{b_5 + b_7 Z_H}{\sqrt{s_{55} + {Z_H}^2 s_{77} + 2 Z_H s_{57}}} \neq 0$
(b)	(1) and (3)	$t = \dfrac{b_4 + b_7 W_H}{\sqrt{s_{44} + {W_H}^2 s_{77} + 2 W_H s_{47}}} \neq 0$
(c)	(2) and (4)	$t = \dfrac{b_4 + b_7 W_L}{\sqrt{s_{44} + {W_L}^2 s_{77} - 2 W_L s_{47}}} \neq 0$
(d)	(3) and (4)	$t = \dfrac{b_5 + b_7 Z_L}{\sqrt{s_{55} + {Z_L}^2 s_{77} - 2 Z_L s_{57}}} \neq 0$
(e)	(1) and (4)	$t = \dfrac{b_4 (Z_H - Z_L) + b_5 (W_H - W_L) + b_7 (Z_H W_H - Z_L W_L)}{\sqrt{\begin{array}{l}(Z_H - Z_L)^2 s_{44} + (W_H - W_L)^2 s_{55} \\ + (Z_H W_H - Z_L W_L)^2 s_{77} \\ +2\left[\begin{array}{l}(Z_H - Z_L)(W_H - W_L) s_{45} \\ +(Z_H - Z_L)(Z_H W_H - Z_L W_L) s_{47} \\ +(W_H - W_L)(Z_H W_H - Z_L W_L) s_{57}\end{array}\right]\end{array}}} \neq 0$
(f)	(2) and (3)	$t = \dfrac{b_4 (Z_H - Z_L) + b_5 (W_L - W_H) + b_7 (Z_H W_L - Z_L W_H)}{\sqrt{\begin{array}{l}(Z_H - Z_L)^2 s_{44} + (W_L - W_H)^2 s_{55} \\ + (Z_H W_L - Z_L W_H)^2 s_{77} \\ +2\left[\begin{array}{l}(Z_H - Z_L)(W_L - W_H) s_{45} \\ +(Z_H - Z_L)(Z_H W_L - Z_L W_H) s_{47} \\ +(W_L - W_H)(Z_H W_L - Z_L W_H) s_{57}\end{array}\right]\end{array}}} \neq 0$

Figure Captions

Figure 1. Moderating effect of autonomy on the training provision-job performance relationship (two-way interaction with continuous moderator).

Figure 2. Moderating effect of autonomy and experience on the training provision-job performance relationship (three-way interaction with continuous moderators).

Figure 3. Examples of quadratic plots

Figure 4. Moderating effect of autonomy on the curvilinear training provision-job performance relationship (two-way quadratic interaction with continuous moderator).

Figure 5. Curvilinear moderating effect of autonomy on the linear training provision-job performance relationship (two-way interaction with quadratic moderator).

Figure 6. Moderating effect of autonomy and experience on the curvilinear training provision-job performance relationship (three-way quadratic interaction with continuous moderators).

Figure 7. Moderating effect of age on the work pressure-absenteeism relationship (two-way binary logistic interaction with continuous moderator).

Figure 8. Moderating effect of age on the work pressure-absence episodes relationship (two-way Poisson interaction with continuous moderator).

Figure 9. Moderating effect of training provision on the autonomy-job performance relationship (a reversal of the plot in figure 1).

Figure 1

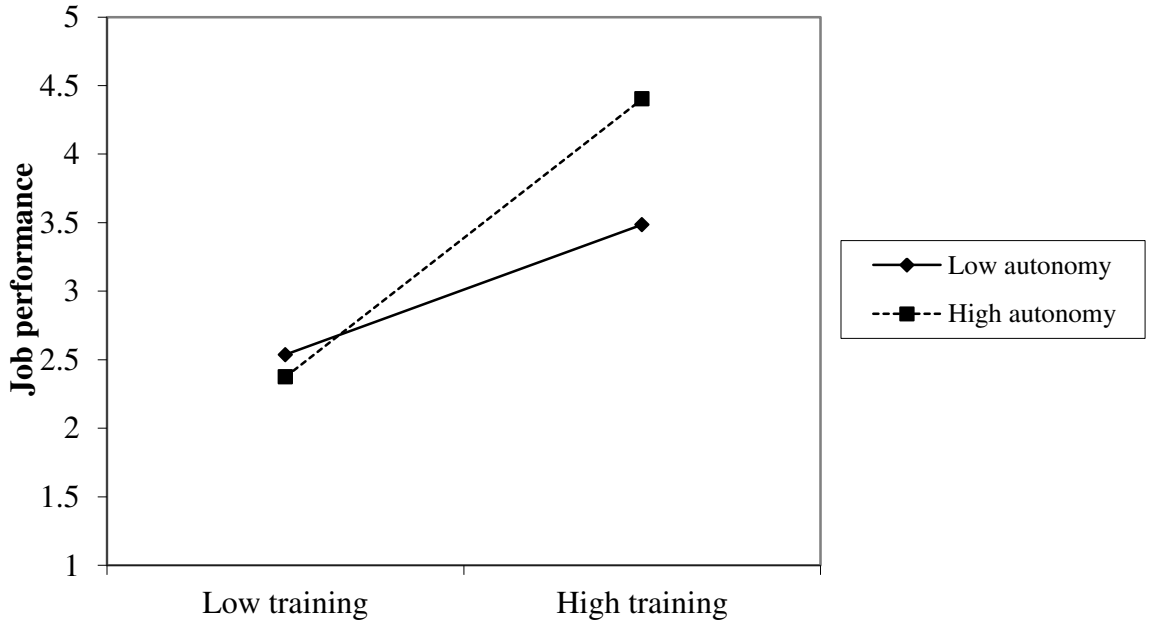

5
4.5
4
3.5
3
2.5
2
1.5
1
Job performance
Low training
High training
Low autonomy
High autonomy

Figure 2

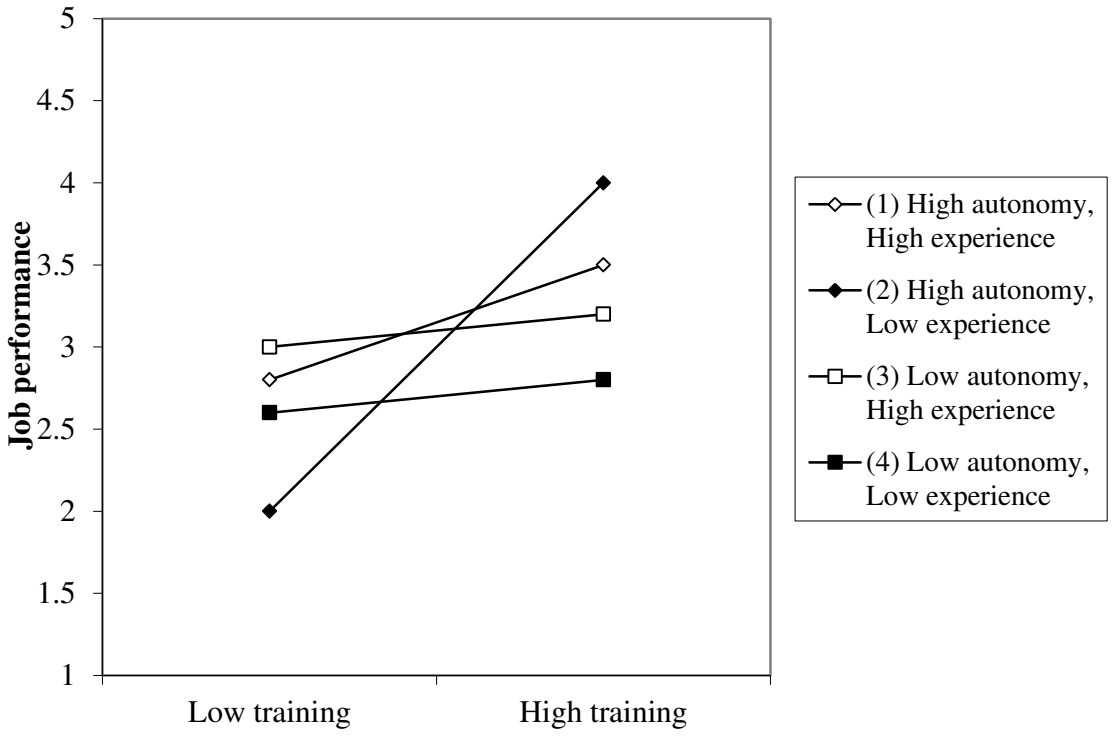

5
4.5
4
3.5
3
2.5
2
1.5
1
Job performance
Low training
High training
(1) High autonomy, High experience
(2) High autonomy, Low experience
(3) Low autonomy, High experience
(4) Low autonomy, Low experience

Figure 3

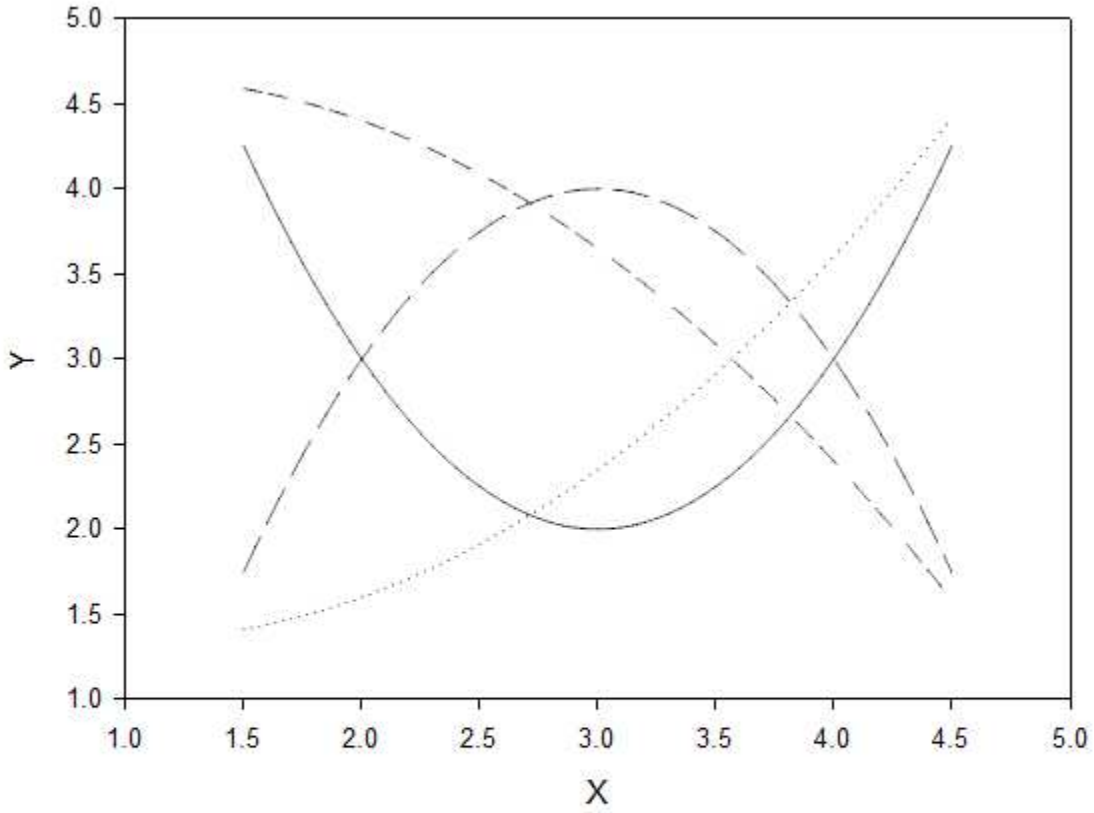

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Y
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
X
b0 = 1, b1 = -6, b2 = 1
b0 = -5, b1 = 6, b2 = -1
b0 = 1.6, b1 = -0.5, b2 = 0.25
b0 = 4.4, b1 = 0.5, b2 = -0.25

Figure 4

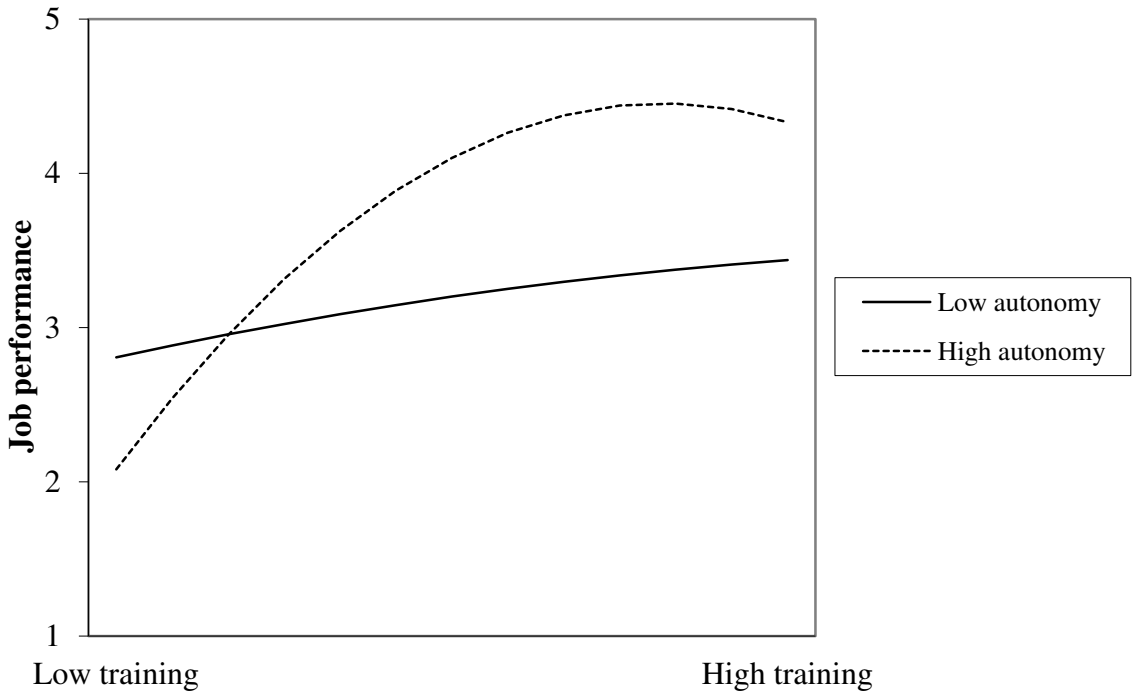

5
4
3
2
1
Job performance
Low training
High training
Low autonomy
High autonomy

Figure 5

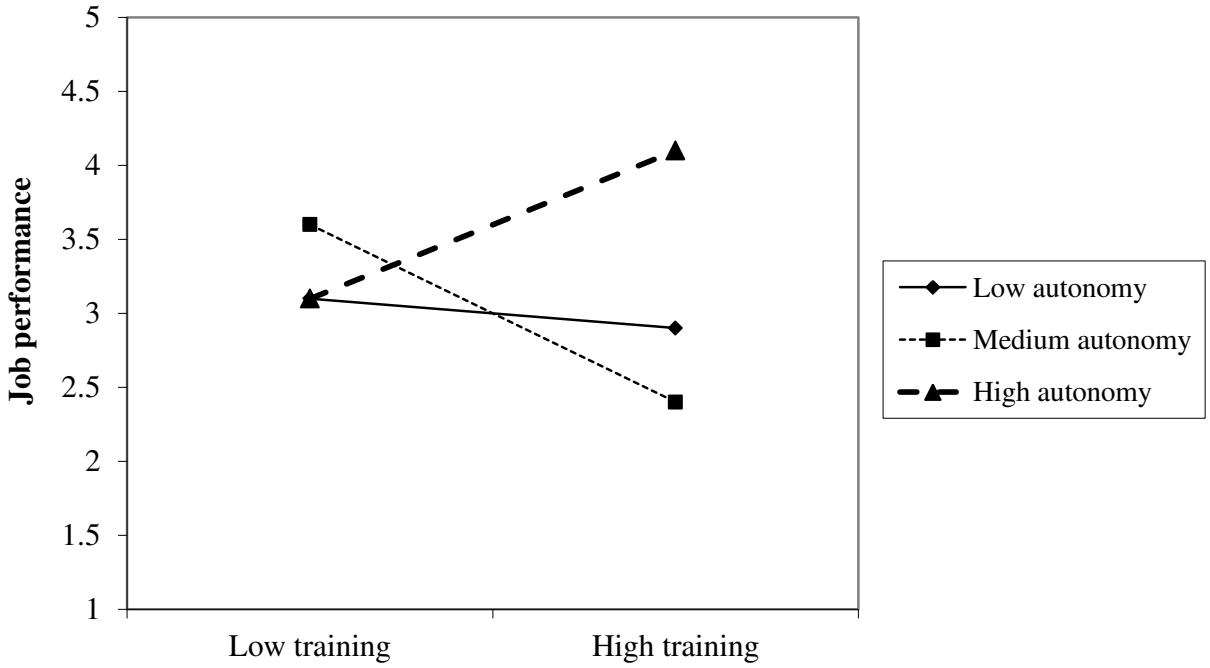

Job performance
5
4.5
4
3.5
3
2.5
2
1.5
1
Low training
High training
Low autonomy
Medium autonomy
High autonomy

Figure 6

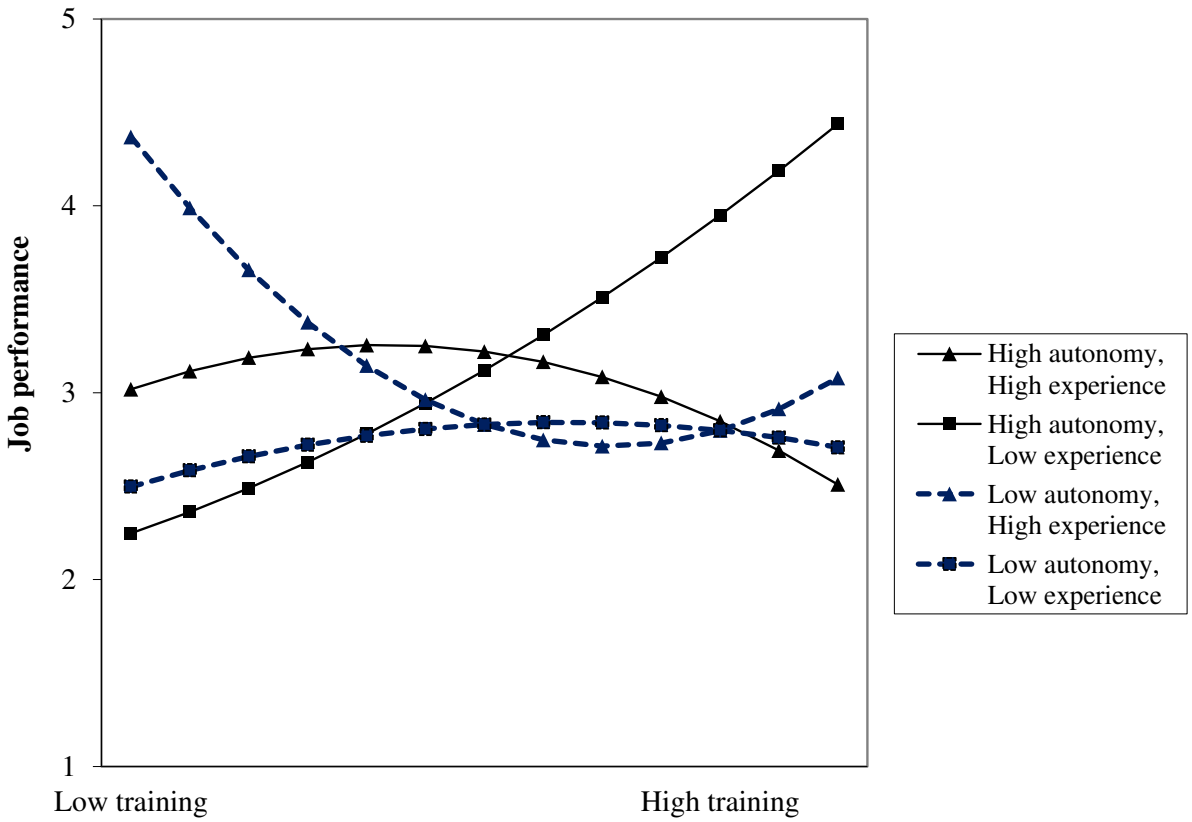

5
4
3
2
1
Job performance
Low training
High training
High autonomy, High experience
High autonomy, Low experience
Low autonomy, High experience
Low autonomy, Low experience

Figure 7

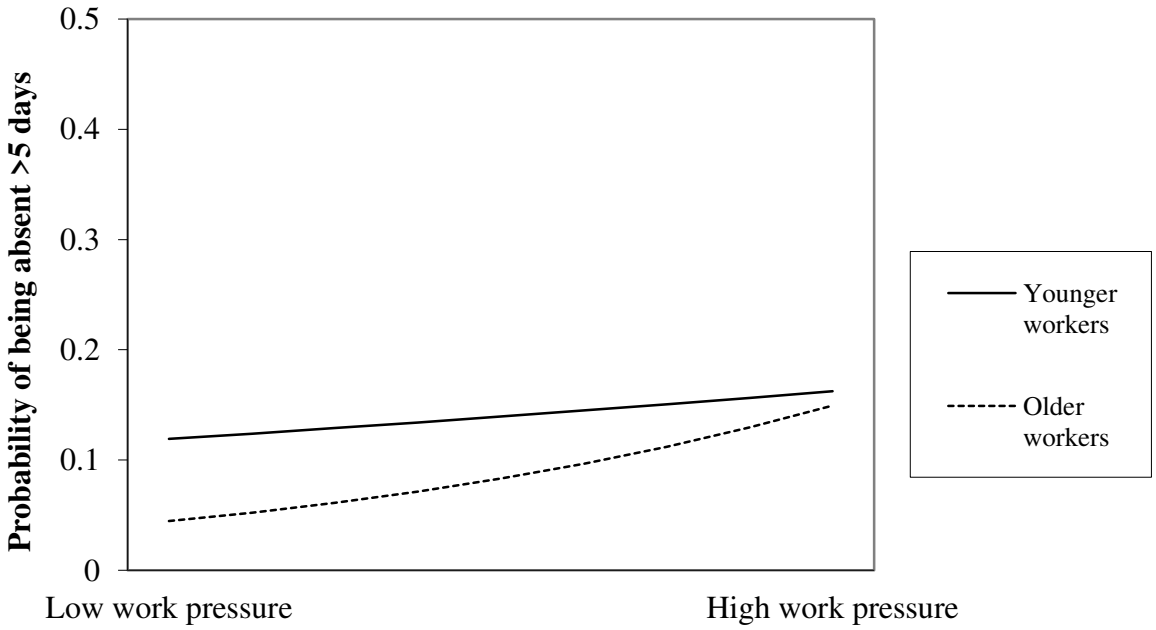

0.5
0.4
0.3
0.2
0.1
0
Probability of being absent >5 days
Low work pressure
High work pressure
Younger workers
Older workers

Figure 8

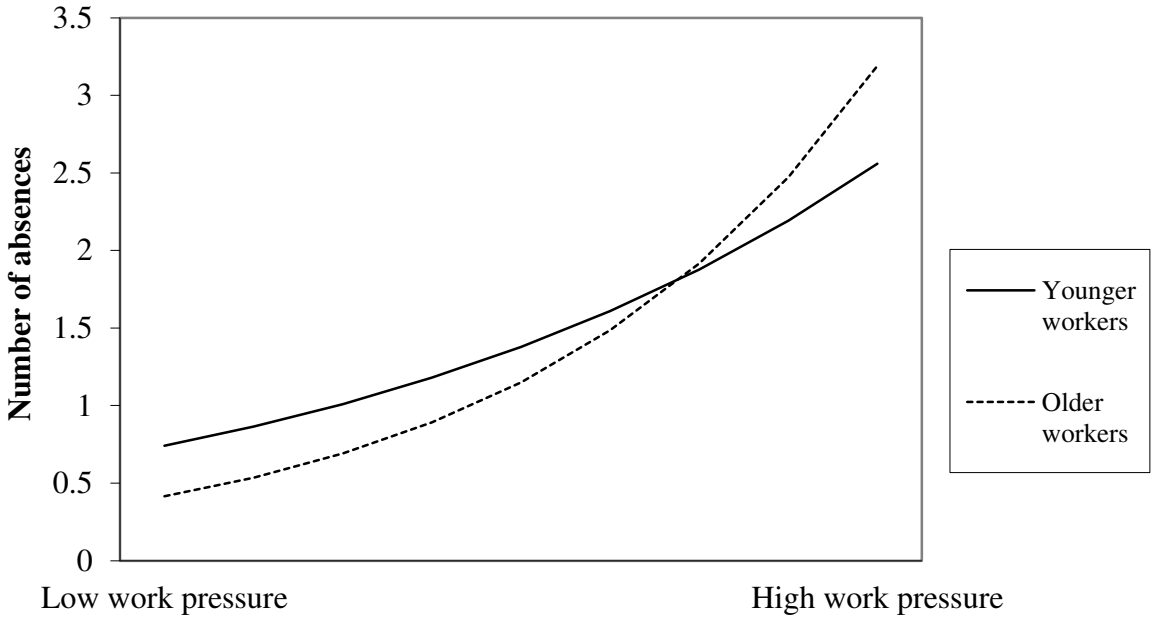
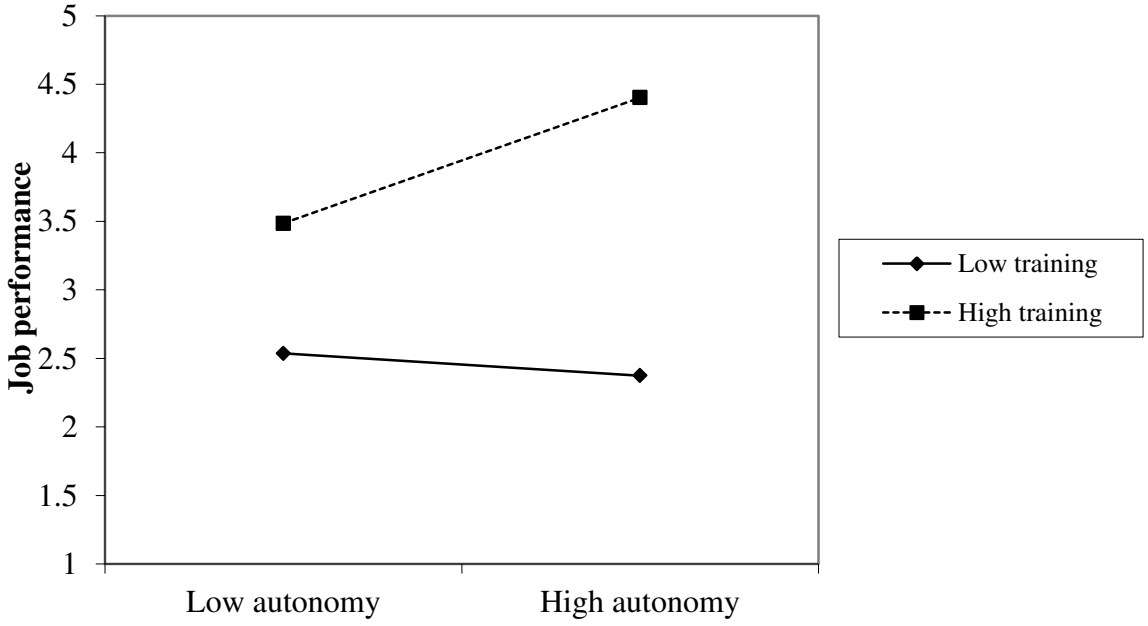

3.5
3
2.5
2
1.5
1
0.5
0
Number of absences
Low work pressure
High work pressure
Younger workers
Older workers

Figure 9


5
4.5
4
3.5
3
2.5
2
1.5
1
Job performance
Low autonomy
High autonomy
Low training
High training

Appendix: SPSS syntax for the models described in the article

In this appendix, variable names are shown in upper case (capitals) and SPSS commands in lower case. In reality the case does not matter for either.

Variable names:
Dependent variable (DV) 1 – PERFORM (Job performance – continuous)
Dependent variable (DV) 2 – ABSENCE (Absenteeism – binary)
Dependent variable (DV) 3 – TIMESABS (Number of occasions absent – discrete)
Independent variable (IV) 1 – TRAIN (Training provision – continuous; mean 3.42)
Independent variable (IV) 2 – WKPRES (Work pressure – continuous; mean 2.95)
Moderator 1 – AUTON (Autonomy – continuous; mean 3.20)
Moderator 2 – EXPER (Experience – continuous; mean 6.5)
Moderator 3 – AGE (Age – continuous; mean 43.7)
Moderator 4 – ROLE (Job role – three levels, scored 1, 2, 3)

* Descriptions of what the commands are doing are shown with asterisks in front of them. SPSS will ignore any commands that begin with an asterisk. Note that the execute commands are not necessary to run any analysis, but commands that do not produce output will not be performed until either a command that does produce output, or an execute command, is run afterwards.

```
* (1) Syntax to create centered versions of continuous IVs and moderators:

compute TRAINC = TRAIN – 3.42.
compute WKPRESC = WKPRES – 2.95.
compute AUTONC = AUTON – 3.20.
compute EXPERC = EXPER – 6.5.
compute AGEC = AGE – 43.7.
execute.


* (2) Syntax to created standardized versions of continuous IVs and moderators. N.B. the variables
     created will have the same names as the originals but preceded by a Z.

descriptives TRAIN WKPRES AUTON EXPER AGE /save.


* (3) Syntax to create and test 2-way interaction with continuous moderator. Note that the inclusion of
     "bcov" ensures that coefficient variances and covariances are included in the output – helpful for
     testing of simple slopes.

compute TRAXAUT = TRAINC*AUTONC.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC AUTONC
 /method = enter TRAXAUT.


* (4) Syntax for alternative method of testing simple slope: for value of AUTON = 4.10. Test of simple
     slope is given by significance of TRAINC term in final model.
```

```
compute AUTONT = AUTON – 4.10.
compute TRAXAUTT = TRAINC*AUTONT.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC AUTONT
 /method = enter TRAXAUTT.
```

* (5) Syntax for ANCOVA version of 2-way interaction with categorical moderator (NB can involve any combination of continuous and categorical IVs/moderators; categorical variables follow "by" command and continuous variables follow "with" command.

```
glm PERFORM by ROLE with TRAINC
 /print = parameter
 /design = TRAINC ROLE TRAINC*ROLE.
```

* (6) Syntax to create and test 3-way interaction with continuous moderators.

```
compute TRAXAUT = TRAINC*AUTONC.
compute TRAXEXP = TRAINC*EXPERC.
compute AUTXEXP = AUTONC*EXPERC.
compute TRXAUXEX = TRAINC*AUTONC*EXPERC.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC AUTONC EXPERC TRAXAUT TRAXEXP AUTXEXP
 /method = enter TRXAUXEX.
```

* (7) Syntax to test curvilinear interaction.

```
compute TRAINSQ = TRAINC*TRAINC.
compute TRASXAUT = TRAINSQ*AUTONC.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC TRAINSQ AUTONC
 /method = enter TRAXAUT TRASXAUT.
```

* (8) Syntax to test for (linear or curvilinear) relationship between IV and DV at particular value of moderator ("simple curve") at AUTON = 4.10. The simple curve or slope is significant if the second step adds significant variance to the model.

```
compute TRAINSQ = TRAINC*TRAINC.
compute AUTONT = AUTON – 4.10.
compute TRAXAUTT = TRAINC*AUTONT.
compute TRASXAUTT = TRAINSQ*AUTONT.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter AUTONT TRAXAUTT TRASXAUTT
```

```
 /method = enter TRAINC TRAINSQ.
```

* (9) Syntax to test for difference between simple curves in a curvilinear three-way interaction. In this example the test is for difference between the curvilinear effect of training provision on job performance at high autonomy/high experience and high autonomy/low experience. High autonomy is defined by AUTON = 4.10. The simple curves are significantly different if the second step adds significant variance to the model.

```
compute TRAINSQ = TRAINC*TRAINC.
compute AUTONT = AUTON – 4.10.
compute TRAXAUTT = TRAINC*AUTONT.
compute TRASXAUTT = TRAINSQ*AUTONT.
compute TRAXEXP = TRAINC*EXPERC.
compute AUTTXEXP = AUTONT*EXPERC.
compute TRXAUTXEX = TRAINC*AUTONT*EXPERC.
compute TRSQXAUTXEX = TRAINSQ*AUTONT*EXPERC.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC TRAINSQ AUTONT EXPERC TRAXEXP AUTTXEXP TRXAUTXEX
TRSQXAUTXEX
 /method = enter TRAXAUTT TRASXAUTT.
```

* (10) Syntax to test 2-way interaction with continuous moderator and binary dependent variable. Note that it is not necessary to create the interaction term separately, but it is still advisable to use centered versions of the IV and moderator.

```
logistic regression variables ABSENCE
 /method = enter WKPRESC AGEC WKPRESC*AGEC.
```

* (11) Syntax to test simple slope of an independent variable (WKPRESC) on binary dependent value (ABSENCE) when the moderator, AGE, is equal to 30. Test of simple slope is given by significance of WKPRESC term.

```
compute AGET = AGE – 30.
logistic regression variables ABSENCE
 /method = enter WKPRESC AGET WKPRESC*AGET.
```

* (12) Syntax to test 2-way interaction with continuous moderator and count dependent variable using Poisson regression. For negative binomial alternative use “distribution = negbin” instead.

```
genlin TIMESABS with WKPRESC AGEC WKPRESC*AGEC
 /model WKPRESC AGEC WKPRESC*AGEC
 distribution = poisson link = log.
```

* (13) Syntax to create dummy variables for a categorical moderator ROLE.

```
recode ROLE (1=1)(2 3=0) into ROLE1.
recode ROLE (2=1)(1 3=0) into ROLE2.
recode ROLE (3=1)(1 2=0) into ROLE3.
execute.
```

* (14) Syntax to create and test 2-way interaction with categorical moderator (with three levels: for more levels, add in dummy variables and interaction terms accordingly).

```
compute TRAXRO1 = TRAINC*ROLE1.
compute TRAXRO2 = TRAINC*ROLE2.
regression /statistics = r coeff cha anova bcov
 /dependent = PERFORM
 /method = enter TRAINC ROLE1 ROLE2
 /method = enter TRAXRO1 TRAXRO2.
```

Author Note

Correspondence concerning this article should be addressed to Jeremy F. Dawson, Management School, University of Sheffield, Sheffield S10 1FL, United Kingdom. Email: j.f.dawson@sheffield.ac.uk

I am grateful to Ron Landis, Scott Tonidandel and Steven Rogelberg for their comments on earlier drafts of this article.