Oswald, F. L., & Putka, D. J. (2016). Statistical methods for big data: A scenic tour. In S. Tonidandel, E. King, & J. Cortina (Eds.), *Big data at work: The data science revolution and organizational psychology* (pp. 43–63). New York: Routledge.
Petty, R., & Cacioppo, J. T. (2012). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer Science & Business Media.
Porras, J. I., & Robertson, P. J. (1993). Organization development: Theory, practice, and research. In M. D. Dunnette & L. M. Hough (Eds.), *The handbook of industrial and organizational psychology* (Vol. 3, pp. 719–822). Palo Alto, CA: Consulting Psychologists Press.
Putka, D. J., & Oswald, F. L. (2016). Implications of the big data movement for the advancement of I-O psychology and practice. In S. Tonidandel, E. King, & J. Cortina (Eds.), *Big data at work: The data science revolution and organizational psychology* (pp. 43–63). New York: Routledge.
Saari, L. M., & Judge, T. A. (2004). Employee attitudes and job satisfaction. *Human Resource Management, 43*(4), 395–407.
Scherbaum, C. A., Putka, D. J., Naidoo, L. J., & Youssefnia, D. (2010). Key driver analyses: Current trends, problems, and alternative approaches. In S. Albrecht (Ed.), *Handbook of employee engagement*. Camberley, UK: Edward-Elgar Publishing House.
Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*(5), 529–540.
Staw, B. M., & Ross, J. (1985). Stability in the midst of change: A dispositional approach to job attitudes. *Journal of Applied Psychology*, *70*(3), 469–480.
Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, *51*(1), 539–570.

Best Practice Recommendations for Conducting Key Driver Analyses

Jeff W. Johnson
CEB

In their critical review of survey key driver analyses (SKDA), Cucina, Walmsley, Gast, Martin, and Curtin (2017) contend that methodological issues limit the usefulness of SKDA and recommend that survey providers stop conducting SKDA until these issues can be overcome. I contend that many of these methodological issues are either overstated or able to be addressed through the proper application of the technique by a competent professional. In this commentary, I make recommendations for how SKDA should be applied so that methodological issues are addressed and the value of SKDA is maximized. Many of these recommendations were made in Lundby and Johnson (2006), who were cited by Cucina et al. but did not have much impact on their focal article.

Jeff W. Johnson, CEB.
Correspondence concerning this article should be addressed to Jeff W. Johnson, CEB, 15 Marcin Hill, Burnsville, MN 55337. E-mail: jejohnson@cebglobal.com

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

The first recommendation is that survey researchers use relative weight analysis (RWA; Johnson, 2000) rather than correlations or multiple regression to identify key drivers. I actually developed RWA for the purpose of identifying key drivers in survey analysis while accounting for the problem of multicollinearity. When predictor variables are correlated, zero-order correlations and standardized regression coefficients have long been considered inadequate (e.g., Darlington, 1968; Green & Tull, 1975; Hoffman, 1960). The primary problem is that these indices often provide nonsensical results such as negative or zero importance for attributes that *should* be important under conditions that are common in survey research (e.g., high variable intercorrelations). A rich literature on determining the relative importance of predictors has developed in statistics, psychology, marketing, economics, and medicine (e.g., Budescu, 1993; Gibson, 1962; Green, Carroll, & DeSarbo, 1978; Healy, 1990; Johnson, 2000; Kruskal, 1987). Cucina et al. virtually ignore this research by focusing their analyses on correlations and stepwise regression. Many of their criticisms of SKDA are actually just criticisms of stepwise regression and do not apply to RWA.

In what follows, I address each of the methodological limitations Cucina et al. identify and their own recommendations for analyzing survey data.

Issue 1: SKDA Results Are Impacted by Item Standard Deviations

In their own survey dataset, Cucina et al. (2017) find a strong positive correlation ($r = .62$) between survey item standard deviations (SDs) and correlations with an overall evaluation item. Although a lack of variance in a variable will limit the size of a correlation, it is unusual for range restriction to differentially influence correlations to such an extent in survey data. I have to wonder how typical this relationship is of most surveys. I performed the same analysis on a 41-item survey with 5,700 respondents and found a correlation between SD and r of .01. The illustrative items Cucina et al. provide may provide a clue as to why there is such a difference. The item with a small SD and small correlation with overall job satisfaction ("Employees use information technology [for example, intranet, shared networks] to perform work") is something you would expect a lot of agreement on but is not something you would expect to influence job satisfaction. By contrast, the item with a large SD and a high correlation ("I have a high level of respect for my organization's senior leaders") is something you would expect little agreement on because respondents are likely to have different opinions on the performance of senior leaders, and those opinions would be expected to influence their overall satisfaction. If the whole survey contains multiple items that have high agreement on things that should have little effect on the criterion, that could explain the strong correlation between SD and r.

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

It would be a mistake for a survey researcher to draw conclusions from only the key driver statistic. It is necessary to examine means and SDs in conjunction with the RWA. An extremely small SD will limit the ability of an item to be identified as a key driver and needs to be considered in the interpretation. I am sure all researchers look at means to determine how the organization is doing on each issue. The best practice is to plot performance (i.e., mean score or percent favorable) against importance (i.e., relative weight) to identify areas on which to take action. A low mean indicates that item represents an issue that should be considered, but it does not necessarily influence the criterion. If the mean is very high, whether that issue is driving the criterion matters little because it does not need to be addressed anyway. Primary attention should be placed on issues that show both low performance and high importance, which requires examining more than just a key driver analysis.

Another issue that Cucina et al. note in this section is the broader nature of scale scores (i.e., composites of items measuring a similar construct) compared to item scores. Using scale scores as predictors would cause a problem for action planning because it would be more difficult to target the appropriate action item when SKDA results are only expressed in terms of a broad scale. This is why I have recommended a simple, two-step approach when applying RWA to a survey consisting of multiple items measuring different topics (Lundby & Johnson, 2006). The first step is to compute scale scores for groups of items measuring a similar topic and use these scale scores as the predictors in RWA. There are several reasons for this, including better reliability for scale scores and keeping the number of predictors to a manageable number.

More important, however, is that relative weights for single items can be misleading if there are differing numbers of items measuring similar topics. Because the importance of highly correlated items is spread out more or less symmetrically among them, a large number of items measuring one topic would tend to get smaller relative weights than a small number of items measuring an equally important topic. As a simplified example, imagine two survey topics, supervision and pay, each of which account for 10% of the predictable variance in the criterion of overall job satisfaction. Suppose supervision is measured by five items, and pay is measured by two items. If a relative weight analysis were run at the item level, that 10% of the predictable variance would be spread out among five items for supervision, so the average relative weight for each item would be 2%. The 10% is spread out among only two items for pay, so the average relative weight for each item would be 5%. Both of the pay items would appear to be more important than any of the supervision items, so the conclusion would be to work only on pay issues. The scale-level analysis shows that supervision is

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

as important as pay, however, so equal emphasis should be given to both topics.

To report on the importance of single items, the scale-level RWA that evaluates the relative importance of general survey topics should be followed up with separate RWAs of the items within each topic. For example, RWA would be conducted on the five items composing the supervision topic to derive relative weights for each of those items. If some aspects of supervision are more important to overall job satisfaction than others, this analysis would identify the specific areas to target for improvement.

Issue 2: SKDA Capitalizes on Chance

Cucina et al. (2017) accuse SKDA of dustbowl empiricism because a priori hypotheses are not made. Actually, RWA and other types of SKDA are data analysis methods, so it is the user that should be accused of dustbowl empiricism, not the tool. There is nothing about SKDA that prevents the user from making hypotheses. I would argue, however, that items are selected to include on a survey because they measure things that are thought to have some relationship with the criterion. For example, a job satisfaction survey includes items that measure different facets of job satisfaction, based on research literature, previous research in a variety of organizations, or research within the client organization (e.g., employee focus groups). In essence, the hypothesis is that all constructs measured by the survey are related to the criterion. The SKDA is conducted to determine the relative strength of these relationships, which is difficult to hypothesize.

Issue 3: Rational or Random Item Selection Performs as Well as SKDA

Cucina et al. (2017) recommend rational item selection in determining key drivers. As just mentioned, a rational approach is typically used to write survey items and all items measuring specific aspects of the general construct to be predicted (e.g., facets of job satisfaction predicting overall job satisfaction) are potential key drivers. Items that do not measure something that is expected to influence the overall evaluation should not be included in the SKDA.

Cucina et al. show that rationally choosing a subset of items yielded nearly as high a validity coefficient as choosing items based on stepwise regression after cross-validating, concluding that rational selection approximated the results of stepwise regression. The fact is that survey items tend to be highly intercorrelated, which is why analysis methods other than multiple regression were needed in the first place (e.g., Green et al., 1978). Because of these high intercorrelations, any group of survey items (even those selected at random) will provide a pretty good multiple correlation with the criterion just because people tend to respond to most items similarly.

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

A high validity coefficient is not evidence for a meaningful set of key drivers. The purpose of SKDA is explanation, not prediction. RWA is intended to explain how variables contribute to the prediction of a criterion, not to select the subset of variables that maximizes prediction (Johnson, 2000). Two items that are highly correlated and have similar correlations with the criterion will have similar relative weights, but the second variable does not improve prediction much beyond that for which the first variable accounts. Demonstrating that two subsets of variables predict the criterion similarly tells us nothing about the similarity in relative importance for those variables.

Issue 4: A Survey's Latent Factor Structure Underlies SKDA Results

Cucina et al. (2017) state that an item could be selected as a key driver because it measures the same factor as the criterion item, which provides an alternative explanation for the correlations between specific items and the criterion. This really is not an issue if we think of the latent structure underlying survey responses as a hierarchical factor model with the overall evaluation at the highest level and the more specific facets beneath it. The criterion item measures the same factors as those measured by the more specific items to the extent that constructs lower in the hierarchy contribute to the construct at the higher level. This is exactly how SKDA should be described from a latent variable perspective.

Issue 5: SKDA Assumes Nonexistent Situational Specificity

Cucina et al. (2017) contend that survey items represent universal constructs and that relationships between specific facets and overall evaluations should be generalizable across situations. They demonstrated similar validity coefficients for key drivers identified by stepwise regression at the government-wide, organizational, and suborganizational levels in their federal survey data. I am not convinced by the argument or the demonstration. Logic alone tells us that there is no reason to expect universality in the relative importance of facets of a construct measured by a survey to an overall evaluation of that construct. Universality requires us to believe that there are no individual differences in the priority people place on different things when making an overall evaluation. The idea that something that is very important to one person could be irrelevant to another is the bedrock concept underlying theories of person–environment fit (Su, Murdock, & Rounds, 2015). For example, Holland's (1997) theory of vocational interests, Dawis and Lofquist's (1984) theory of work adjustment, and Schneider's (1987) attraction-selection-attrition theory are all based on the idea that there are individual differences in work preferences and the importance placed on elements of the job and organization. SKDA seeks to identify generalities

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

across individuals within an organization to target the issues that will have the greatest impact on the most people. What those issues are depends on the configuration of individuals within the organization.

Cucina et al. cite validity generalization of tests predicting job performance as a parallel to their argument for the generalizability of key drivers across organizations, but recent research suggests that validity generalization is likely overstated in a variety of domains (Tett, Hundley, & Christiansen, in press). Regardless, any cross-situational consistency in prediction of job performance from psychological assessments tells us nothing about relationships between specific attitudes and general attitudes toward a job or organization.

Cucina et al. acknowledge that finding similar results across organizations may be due to a lack of diversity across organizations in the federal government. This is likely, because the nature of the organization and the kind of people it attracts will surely influence the drivers of satisfaction. For example, satisfaction with pay will probably be a more important driver of satisfaction in a Wall Street brokerage firm than in a nonprofit charitable foundation. Satisfaction with training should be a more important driver in an organization that emphasizes training over selection than in an organization that selects employees based on their ability to already do the job. Federal agencies are pretty homogeneous, so it would be less likely to find these kinds of differences in that sample.

Demonstrating similar validities at different levels for drivers selected by stepwise regression is irrelevant. Again, the level of prediction tells us nothing about the relative importance of the variables. There is no indication that the same items were selected at each level or that the relative importance of the items selected at each level was the same.

Issue 6: SKDA Assumes Nonexistent Temporal Specificity

I agree that SKDA is not necessary each year a survey is administered, but that is a decision made by the researcher, not an issue with the analysis technique. There is nothing about SKDA that demands the analysis be repeated with each survey administration. A good reason to conduct a new SKDA would be if a new topic is added to the survey and the relative importance of that topic compared to the others is of interest, or perhaps there has been a significant organizational change that would suggest a shift in the relative importance of certain survey topics.

Issue 7: Multivariate Analyses Are Suspect When Variables Outnumber Observations

Cucina et al. (2017) point out the problems associated with conducting a multivariate analysis with a small sample size, and I agree that it makes no

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

sense to conduct RWA at the level of a typically sized team. The sample size for conducting RWA should not be smaller than that appropriate for conducting ordinary multiple regression, but there is no minimum sample size appropriate in all situations (Maxwell, 2000). I recommend a minimum sample size of about 20 times the number of variables and that standard errors around relative weights and differences between relative weights (Johnson, 2004) be computed to give confidence in conclusions.

Issue 8: Correlation ≠ Causation

SKDA uses correlational analysis to derive causal conclusions, so there is always an inferential leap required when labeling a variable a "driver" of something. Certainly it is possible that being generally satisfied overall could cause people to respond favorably to specific items or that a third variable (e.g., positive/negative affectivity) causes responses to both kinds of items. Nevertheless, it is reasonable to think that a person's overall evaluation of a construct would be driven by his or her opinions with respect to specific aspects of that construct.

Cucina et al. recommend conducting experiments with random assignment of subjects to conditions in organizational settings to establish causation. Although this would be ideal, it is not realistic. It is neither possible nor ethical to randomly assign employees to conditions such that you control their experiences with job security, pay, the supervisor, or interactions with coworkers. Applied research has limitations that often require the researcher to make some assumptions. This is true of many programs of research in I-O psychology, not just SKDA.

Conclusion

Cucina et al. (2017) conclude by proposing a new four-step approach to survey interventions, but the only step that would replace SKDA is Step 1: Identify targets for intervention. That is what SKDA does, so the remaining steps can still be applied regardless of how targets for intervention are identified. Cucina et al.'s suggestion to replace SKDA is essentially to focus on the areas that have the lowest scores relative to a norm group, but that is the approach that has been applied for years and the reason why SKDA has become so popular. Intervening in areas that have low scores will not impact the criterion if respondents tend not to see those areas as important to their overall evaluation. SKDA offers the ability to combine performance and importance information so that interventions can be targeted to areas with low performance and high importance.

SKDA is a tool. Many of the criticisms Cucina et al. offer of SKDA seem to be criticisms of the potential application of the tool rather than of the tool

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press

itself. In their article, criticisms specific to the tool are primarily directed at stepwise regression, which has long been discredited as a measure of predictor variable relative importance. Therefore, I recommended the use of RWA rather than correlation or regression to conduct SKDA and present recommendations for how to apply it when analyzing survey data. Following these suggestions makes a moratorium on SKDA unnecessary and will allow survey researchers to better serve their clients.

References

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542–551.

Cucina, J. M., Walmsley, P. T., Gast, I. F., Martin, N. R., & Curtin, P. (2017). Survey key driver analysis: Are we driving down the right road? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 10*(2), 234–257.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*, 161–182.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis: University of Minnesota Press.

Gibson, W. A. (1962). Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy. *Psychological Reports*, *11*, 32–34.

Green, P. E., & Tull, D. S. (1975). *Research for marketing decisions* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Green, P. E., Carroll, J. D., & DeSarbo, W. S. (1978). A new measure of predictor variable importance in multiple regression. *Journal of Marketing Research*, *15*, 356–360.

Healy, M. J. R. (1990). Measuring importance. *Statistics in Medicine*, *9*, 633–637.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, *57*, 116–131.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1–19.

Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods*, *7*, 283–299.

Kruskal, W. (1987). Relative importance by averaging over orderings. *American Statistician*, *41*, 6–10.

Lundby, K. M., & Johnson, J. W. (2006). Relative weights of predictors: What is important when many forces are operating. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, methods, and applications* (pp. 326–351). San Francisco, CA: Jossey-Bass.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, *5*, 434–458.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*, 437–453.

Su, R., Murdock, C., & Rounds, J. (2015). Person-environment fit. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career intervention: Vol. 1. Foundations* (pp. 81–98). Washington, DC: American Psychological Association.

Tett, R. P., Hundley, N., & Christiansen, N. D. (in press). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology: Perspectives on Science and Practice*. doi: 10.1017/iop.2017.26

https://doi.org/10.1017/iop.2017.22 Published online by Cambridge University Press