Industrial and Organizational Psychology, 10(2), pp 234–257 June 2017.
Copyright © 2017 Society for Industrial and Organizational Psychology. doi:10.1017/iop.2016.97

Focal Article

Survey Key Driver Analysis: Are We Driving Down the Right Road?

Jeffrey M. Cucina, Philip T. Walmsley, and Ilene F. Gast
U.S. Customs and Border Protection

Nicholas R. Martin
Aon Consulting

Patrick Curtin
National Science Foundation

One of the typical roles of industrial–organizational (I-O) psychologists working as practitioners is administering employee surveys measuring job satisfaction/engagement. Traditionally, this work has involved developing (or choosing) the items for the survey, administering the items to employees, analyzing the data, and providing stakeholders with summary results (e.g., percentages of positive responses, item means). In recent years, I-O psychologists moved into uncharted territory via the use of survey key driver analysis (SKDA), which aims to identify the most critical items in a survey for action planning purposes. Typically, this analysis involves correlating (or regressing) a self-report criterion item (e.g., "considering everything, how satisfied are you with your job") with (or on) each of the remaining survey items in an attempt to identify which items are "driving" job satisfaction/engagement. It is also possible to use an index score (i.e., a scale score formed from several items) as the criterion instead of a single item. That the criterion measure (regardless of being a single item or an index) is internal to the survey from which predictors are drawn distinguishes this practice from linkage research. This methodology is not widely covered in survey methodology coursework, and there are few peer-reviewed articles on it. Yet, a number of practitioners are marketing this service to their clients. In this focal article, a group of practitioners with extensive applied survey research experience uncovers several methodological issues with SKDA. Data from a large multiorganizational survey are used to back up claims about these issues. One issue

Jeffrey M. Cucina, Philip T. Walmsley, and Ilene F. Gast, U.S. Customs and Border Protection, Washington, DC; Nicholas R. Martin, Aon Consulting, Washington, DC; Patrick Curtin, National Science Foundation, Arlington, Virginia.

Ilene F. Gast is now retired.

The views expressed in this article are those of the authors and do not necessarily reflect the views of U.S. Customs and Border Protection, the National Science Foundation, or the U.S. federal government. Portions of this article were presented at the 2011 and 2012 meetings of the Society for Industrial and Organizational Psychology and the January 2013 meeting of the U.S. Office of Personnel Management's FedPsych Forum.

Correspondence concerning this article should be addressed to Jeffrey M. Cucina, 1400 L Street, NW, 7th Floor, Washington, DC 20229–1145. E-mail: jcucina@gmail.com

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

is that SKDA ignores the psychometric reality that item standard deviations impact which items are chosen as drivers. Another issue is that the analysis ignores the factor structure of survey item responses. Furthermore, conducting this analysis each time a survey is administered conflicts with the lack of situational and temporal specificity. Additionally, it is problematic to imply causal relationships from the correlational data seen in most surveys. Most surprisingly, randomly choosing items out of a hat yields validities similar to those from conducting the analysis. Thus, we recommend that survey providers stop conducting SKDA until they can produce science that backs up this practice. These issues, in concert with the lack of literature examining the practice, make rigorous evaluations of SKDA a timely inquiry.

The past decade has seen an increase in the popularity of "survey key driver analysis" (SKDA) to help decision makers focus their efforts on the most critical organizational issues (Colquitt & Macey, 2005; Kraut, 2006; Rasch, Kowske, & Herman, 2009). This analysis attempts to identify a set of survey items (i.e., drivers) that have the greatest impact on a specified organizational outcome (e.g., overall job satisfaction). Typically, scores on a single, global item (e.g., overall job satisfaction) are correlated with or regressed onto the scores for the remaining items within a survey. However, it is also possible to use an index score, which is a scale score formed from multiple items, as the criterion. The items with the strongest relationship are labeled as "key drivers" of the construct measured by the global item, and organizational leaders are encouraged to focus their efforts and resources on the "key drivers." The use of criteria internal to the same survey as the items used as predictors distinguishes SDKA from linkage analysis (Wiley, 1996; Wiley & Campbell, 2006) with which some industrial–organizational (I-O) psychologists are familiar.

Increasingly, SKDA has become part of mainstream practice. A manual search of advertisers in *The Industrial Psychologist* yielded 10 vendors who are now offering key SKDA as one of their services, with statements that this methodology is "a traditional approach" and a "best practice." Despite the recent prevalence of SKDA, there is a dearth of empirical literature on the topic. Only a handful of presentations, articles, and chapters discuss SKDA. We view this as potentially problematic in that a practice is being implemented about which we know very little from a research and theoretical perspective. Contrast this with the wide body of research and theory that exists for psychometrics and test development.

SKDA Defined

Work by Kraut (2006), Lundby and Johnson (2006), Rasch et al. (2009), and Leonard and Lieberman (2006) provides the framework and reasoning behind SKDA. For the purposes of our review, we are focused on these authors'

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

operationalization of SKDA. This analysis identifies a set of survey items that has the greatest impact on either a single outcome item or an index score (computed from several outcome items). The criterion score represents a key indicator of organizational effectiveness such as overall job satisfaction or turnover intention. The set of items is seen as "driving" the results for the outcome measure and is meant to provide organizational stakeholders a means of focusing limited resources on issues that will yield the greatest improvements in organizational effectiveness. Survey key driver items are identified through analyses based on the general linear model, such as correlation coefficients (including zero-order, partial, or semipartial correlation coefficients) and regression analyses (including stepwise and multiple regression or relative weights analysis). Individual employees' responses are used as the cases, or unit of analysis. A unique set of survey key driver items is computed for each organizational entity, with the implicit assumption that drivers are specific to each entity. Although SKDA could be conducted using scale scores as predictors, we focus our focal article on the use of item scores as predictors as this approach is more prevalent. We also focus on the use of a single criterion item, as it is what we have observed most frequently in practice.

Our Position: Methodological Issues Limit Usefulness of SKDA

In this focal article, we take the stance that methodological issues limit the usefulness of SKDA in applied settings. We base our position, in part, on the psychometric literature. We begin with a short justification for the viewpoint that surveys are psychometric instruments and an overview of the methodological issues.

Surveys as Psychometric Instruments

Some readers might question whether organizational surveys can be viewed psychometrically. We would counter that many organizational surveys aim to measure job satisfaction and employee engagement[1] (Macey & Schneider, 2008), which are psychological constructs. Even those not developed with the explicit aim of measuring psychological constructs have psychometric properties. Indeed, Gast (2008, 2010) has shown how psychometric theory can be applied to the development of organizational surveys just as it has traditionally been applied to personnel selection. Both processes begin with the operational definition of constructs and employ judicious item-writing procedures. During implementation, both require knowledge of research design

[1] The distinction between job satisfaction and employee engagement has been debated (see, for example, the inaugural issue of this journal). It is not our intent to resurrect this debate in our focal article; therefore, we refer to both job satisfaction and engagement throughout this article.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

and rely on random sampling for creating psychometric norms. Finally, in the data analysis and reporting phases, both rely on similar descriptive and inferential statistics for scale development and scoring. Several established psychometric concepts appear relevant to SKDA (e.g., item variance, range restriction, latent variables, and factor analysis).

Methodological Issues

We outline several methodological issues with SKDA. The issues cover topics such as the standard deviation (*SD*) of survey items, capitalization on chance, empirical versus rational item selection, the factor structure of organizational surveys, situational and temporal specificity of SKDA, multivariate analysis requirements, and finally the classic issue of correlation and causation. These issues are largely based on psychometric theory; however, we provide empirical support for many of these issues using a publicly available dataset from a survey used by the U.S. federal government. We do this to provide evidence that the issues are not merely conceptual or hypothetical but rather truly impact the interpretation of SKDA results. Information on the survey data we used is provided in Appendix A.

Issue 1: SKDA Results Are Impacted by Item *SD*s

According to psychometric theory, an item's *SD* affects the magnitude of its item-total correlation. Very easy or difficult items (i.e., those with very low or high *p*-values or means) have low *SD*s and cannot have a high item-total correlation. Items with moderate *p*-values or means have the largest *SD*s and highest correlations with the total score (assuming that they measure the same construct as the total score). Similarly, range restriction on a predictor reduces its correlation with a criterion. In SKDA, the items serve as predictors, and psychometric theory predicts that items with lower *SD*s will correlate less with the criterion. If this is true, then there will be a positive correlation between item *SD*s and survey key driver metrics.

Indeed, using the federal survey dataset described in Appendix A, we found a strong positive correlation ($r = .62$) between the *SD* of an item and the bivariate correlation of that item with the criterion item. To illustrate, item 54 ("Employees use information technology [for example, intranet, shared networks] to perform work") had the lowest *SD* (.79) and a correlation of .33 with the criterion item. In contrast, item 37 ("I have a high level of respect for my organization's senior leaders") had the highest *SD* (1.22) and a correlation of .70 (about twice as strong) with the criterion item. As mentioned earlier, it is possible to conduct SKDA using an index score as the criterion. We created an index score using three criterion items (as described in Appendix A) and again found a strong positive correlation

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

($r = .59$) between the *SD* of an item and the bivariate correlation of that item with the criterion index.

The .62 value was obtained using the entire dataset. Most SKDA statistics are obtained separately for organizations and suborganizations. However, when we analyzed the data within organizations and suborganizations, we obtained similar results ($r = .64$ and .68, respectively). Scherbaum, Putka, Naidoo, and Youssefnia (2010) have suggested using relative weights analysis (RWA), a more modern technique that was originally developed for other applications (Johnson, 2000; Johnson & LeBreton, 2004) but aims to address limitations of the bivariate and regression-based approaches. Using RWA, we again found a "high" (as defined by Cohen's, 1992, cutoffs) correlation between item *SD*s and the SKDA metrics: $r = .51$ using the entire dataset. Overall, when examined at the bivariate level, SKDA metrics are highly correlated with item *SD*s. A 1 *SD* increase in the *SD* is associated with an increase in the SKDA metrics of .51 to .68, depending on the metric used and the data analyzed.

We did find that using multivariate regression analysis reduces the correlation between SKDA statistics and item *SD*s. For example, when all of the items are entered into a multiple regression equation, the correlation between the β-weights and item *SD*s was .26. When the items were entered into a stepwise regression procedure that selected five or 10 items, the Nagelkerke correlations between each item's status as a driver (coded as 0 or 1) and item *SD*s were .27 (in both instances). These correlations are better than those from the bivariate analysis, although they come with the caveat that multicollinearity and capitalization on chance negatively impact the interpretability of these β-weights. Additionally, the factor structure underlying an organizational survey might explain the reduced correlation for the SKDA metrics obtained using multivariate analyses. Controlling for other items (via multiple regression) partials out common factors that underlie item responses. These factors may be the underlying dimension that correlates with item *SD*s. Interestingly, partial and semipartial correlations are not used in SKDA but are well grounded in psychometric theory (Nunnally & Bernstein, 1994). These statistics have a similar intent to RWA but have reduced correlations with item *SD*s to .24 in both cases.

An anonymous reviewer suggested another approach to SKDA, in which scale scores might be used as the predictors rather than individual items. The best fitting factor structure we found suggests that six scales can be formed using the data we studied. Studying the effects of *SD*s of the seven scales on SKDA metrics is problematic with such a small sample size ($n = 7$ scales). Thus, it remains to be seen if *SD*s impact SKDA when scale scores are used (we also note the concomitant issue of *SD*s increasing by virtue of composite variance increasing as items are added, which is problematic when the scale

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

lengths are different). Additionally, scales are by definition broader in nature in terms of topic and construct than individual items, which might make it more difficult for action planning to be conducted. In other words, it is likely more difficult to create an action plan for a group of approximately 10 items (measuring a collection of related but distinct topics) than for a single item. Although the impact of *SD*s on SKDA and scale score usage needs further investigation, the likely intervention development tactic, action planning, could very likely be undermined due to an inability to properly target the appropriate issues within the organization.

In our view, the correlation of SKDA metrics and item *SD*s exposes a major conceptual flaw with SKDA: Items that most employees identify as problem areas for the organization are less likely to be identified as drivers. Consider an organizational unit in which all respondents strongly disagree with a positively worded item. For the sake of argument, let us assume that in reality this item causes the criterion. Responses to this item have zero variance and, therefore, mathematically cannot covary or correlate with any criterion items. Thus, this item, despite its organizational relevance, can never be identified as a driver.

Issue 2: SKDA Capitalizes on Chance

SKDA examines the relationship between multiple items with a criterion item and then selects those items that best predict the criterion item. Essentially SKDA resorts to dustbowl empiricism. In our experience, a priori hypotheses are not made before the SKDA is conducted. Furthermore, sometimes SKDA uses stepwise regression, a practice that is strongly criticized in psychometrics literature because its true Type I error rate grossly exceeds its nominal Type I error rate (Cureton, 1950; Nunnally & Bernstein, 1994). Psychometricians emphasize the importance of cross-validation whenever stepwise regression is conducted; yet, SKDA rarely takes this into account. Using the federal dataset, we found that shrinkage occurs when stepwise regression is used to select groups of five or 10 items, mainly when the data come from small organizations (defined as those organizations whose sample sizes were in the bottom third). After triple cross-validation, the validity coefficients for the predicted criterion scores dropped slightly from .81 to .74 when five items were selected and from .83 to .74 when 10 items were selected. The median sample size for the small organizations was 31; we suspect that shrinkage would be even higher at the work-unit level.

Issue 3: Rational and Even Random Item Selection Performs as Well as SKDA

Even when stepwise regression is not used, SKDA relies on a dustbowl empirical approach to selecting item as drivers; no a priori hypotheses are made concerning which items will or will not predict the criterion items. In

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

contrast, psychometric theorists often recommend a rational approach to item development, taking into account past research on constructs measured by the items and the criterion. Even in instances where empirical methods are used in psychometrics and personnel selection, the importance of a priori and rational hypotheses regarding items is paramount. Church and Waclawski (1998) made similar arguments in the context of linkage analysis (which is similar to SKDA except that an external criterion, such as turnover or profit, is used). They reinforce the importance of "logically" linking survey items to the criterion and recommend "a thorough review of the relevant literature regarding the variables to be linked and a solid research design" (p. 269). These arguments suggest that a rational approach to identifying survey key driver items is best.

We compared the validity of empirical and rational SKDA using the federal survey dataset. Three of the authors rationally rated each item with respect to its relationship with the criterion item and whether change can be actionable at the suborganization, organizational, or governmentwide level. The consistency of the ratings was measured using the intraclass correlation (3, *k*); a value of .611 (95% CI: .517 to .689) was obtained. We chose the five and 10 best items based on the rational key and compared it with the five and 10 item sets from the stepwise regression analysis. The five-item stepwise regression empirical method yielded a triple cross-validity of .83, which was only slightly higher than that from the rational approach ($r = .79$). When 10 items were selected, stepwise regression yielded a triple cross-validity of .82, compared with a validity of .81 from the rational key. Thus, rational ratings reasonably approximated the results of SKDA using stepwise regression.

In the spirit of Cureton's (1950) classic article "Validity, Reliability, and Baloney," we also came up with a rather bold idea for an analysis that yielded a shocking result. We selected five and 10 items from the federal survey using a random number generator and computed the validities. Selecting five and 10 items at random yielded validities (.74, in both cases) that were only slightly lower than empirical SKDA and that were similar to rational SKDA.

The success of empirical item selection is dependent on sample size, and many practitioners are in a position to conduct SKDA for small organizational units. Thus, we replicated this analysis using data from small organizational units, which were defined in the section covering Issue 2. After cross-validation, randomly selected items had the same validity as empirically selected ones when data from small organizational units were used. Thus, when stepwise regression is used, SKDA is little more effective than pulling the items out of a hat, especially for smaller samples. The same can be said for the correlational SKDA methodology. When we selected the items with the top five and 10 correlations with the criterion, the resulting criterion-related validities of the unit-weighted sum of these items were only slightly

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

higher (the largest difference was .10 correlational units) than randomly chosen items.

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

Considerations of the possible factor structure of an organizational survey are often missing from SKDA practice. Factor analysis has an important role in psychometrics, especially in examining the internal structure of an instrument. Psychometric theory posits that latent variables (or factors) underlie responses to items in an inventory (e.g., a test or employee survey). Confirmatory factor analysis (CFA) provides a method for testing the degree to which a given model specifying the underlying or latent variables for item responses explains the pattern of intercorrelations among items. Because employee surveys measure attitudinal constructs, it seems likely that they capture variance attributable to latent variables and that those latent variables should explain correlations between items. As mentioned earlier, SKDA involves correlating a set of predictor items with a criterion item. Psychometric theory predicts that at least some of the correlation between each predictor and the criterion item can be explained by a latent variable underlying the two items.

We found empirical evidence that the latent factor structure underlies SKDA results. We conducted a CFA on the governmentwide data based on the factor model validated by Cucina, Credé, Curtin, Walmsley, and Martin (2014), with the addition of a new factor covering 11 employee benefit items not studied by Cucina et al. (2014). This model yielded acceptable fit statistics (normed fit index = .989; $\chi^2 = 1{,}327{,}518$; $df = 2{,}387$), suggesting that seven oblique factors underlie responses to the survey. We then examined the CFA model-implied correlations between each item and the criterion and the observed correlations. The correlations were strikingly similar; the difference between the correlations was small (mean difference $<.01$, median $<.01$, $SD = .03$, minimum $= -.06$, maximum $= .06$). Overall, the latent factor structure underlying the survey provides an alternative explanation for the correlations obtained using SKDA. An item may be identified as a driver not because it is driving responses to the criterion item but instead because it measures the same factor(s) as the criterion item.

Earlier we noted that SKDA metrics are biased by the *SD* of each item. This effect appears to have independence from that of the factor structure. We found that the factor structure and each item's *SD* independently predicted each item's observed correlation with the criterion. This was done by conducting a multiple regression analysis using the items as cases, each item's *SD* and factor loading as predictors, and each item's observed correlation with the criterion item as the criterion. For the bivariate SKDA, both the item's *SD* and its factor loading had significant β-weights, and the

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

regression equation had a large effect size ($R = .80$, $p < .01$; $\beta_{loading} = .64$, $p < .01$; $\beta_{SD} = .24$, $p = .01$). This implies that the correlations between predictor items and the criterion item can involve two sources of variance: the *SD* of the predictor item and the factor loading. Thus, whether or not an item is identified as a driver (using the correlational SKDA methodology) depends on the item's *SD* and the item's factor loading.

As suggested by an anonymous reviewer, the underlying factor structure of a survey can lead to illusory evidence that addressing a driver item increases job satisfaction/engagement. Recall that the criterion item and the predictor items measure the same latent variable(s)/factor(s). Suppose that an organization identifies an item as a driver, conducts action planning to address that driver, and then observes an increase in job satisfaction/engagement with the next survey administration. It is possible that a third variable, a statistical effect such as regression to the mean, or an organizational behavior phenomenon such as the Hawthorne effect led to an increase in the underlying latent variable(s)/factor(s), which in turn caused observed scores on the predictor and criterion items to increase together. Regardless, one cannot infer that an increase in job satisfaction/engagement has been caused by an intervention designed to address a driver item (as we discuss later in this article).

Issue 5: SKDA Assumes Nonexistent Situational Specificity

SKDA is often performed for each organization that administers a survey; this analysis is performed even when identical item content is used across the different organizations. Many survey vendors now include SKDA as part of their default analysis packages for surveys, conducting them for each organization that completes the survey. This practice implies that the analysis is situationally specific, and that survey key driver items would be expected to vary across organizations. In essence, differences among sets of survey key driver items could be attributed to organizational variables, necessitating the use of SKDA for each organization that takes a survey.

This practice seems at odds with what we have learned from the psychometrics literature. A psychometric approach would posit that the survey items represent universal constructs, such as job satisfaction, and that the relationships between potential driver items and the criterion item should be generalizable across organizations. Although the scores on the survey items may vary across organizations, the construct measured by the items should be constant. Our argument parallels Schmidt and Hunter's (1977) validity generalization (VG), whereby the validity of a test is not situationally specific and differences in validity coefficients are due to sampling error or established moderators. Note also that under VG, the validity of a test can

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

be invariant across different situations; however, the mean test and criterion scores may differ significantly.

Indeed, when we examined the federal survey data, we found a lack of situational specificity. We compared the adjusted R and cross-validities of predicted values from suborganization, organization, and governmentwide regression equations. These results indicated that survey key driver items created at the governmentwide level have very similar validity ($r = .84$) to those created at the organization ($r = .85$) and suborganization levels ($r = .85$). Similar results were obtained when we computed the triple cross-validities for the stepwise regression SKDAs. The results for the five-item stepwise regressions were .83, .83, and .82 for the governmentwide, organization, and suborganization analyses. For the 10-item stepwise regressions, the triple cross-validities were .82, .82, and .81, respectively.

Further analyses indicated that some items are frequently identified as drivers almost regardless of the type and method of analysis used. Examples of the common driver items include "My work gives me a feeling of personal accomplishment," "How satisfied are you with your involvement in decisions that affect your work?" and "Considering everything, how satisfied are you with your pay?" Overall, we found that the practice of conducting SKDA for different organizations is not empirically supported. Instead, it appears to be a waste of resources both on the side of those conducting the analysis and the side of those sponsoring the analysis (in terms of cost and time spent interpreting the results). If a survey vendor or provider wishes to use SKDA, the best approach would be to conduct the analysis one time using all available data, rather than for each client organization's administration. Although there were differences in the mean scores on the survey items across different organizations and suborganizations, the SKDA results tended to be quite similar. That said, a potential caveat to our findings is the use of data from only the U.S. federal government. Data from a more diverse database (e.g., survey responses from multiple countries and public and private sector organizations) might yield different findings.

Issue 6: SKDA Assumes Nonexistent Temporal Specificity

An SKDA is often performed each time a survey is administered to a single organization, suggesting that survey key driver items are temporally specific (i.e., not stable across time). Temporal specificity of survey key driver items is essentially an extension of the situational specificity issue described above; however, years are used as situations, whereas organizations were used as situations in the previous question. There is an intuitive appeal to temporal specificity as employee surveys measures attitudes and opinions, which may change over time.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Looking at the federal survey data, we did not find evidence of temporal specificity. We obtained survey responses for four years of survey administrations and conducted SKDA for each year. We then placed the key driver metrics into a new dataset and examined the interrelationships among the survey key drivers from year to year. At the governmentwide level, the bivariate and β-weight regression survey key driver metrics correlated across years at values approaching unity ($rs = .99$ and .98, respectively). In addition, we computed correlation coefficients between the dichotomous status (0, 1) of each item with respect to the top five and 10 items, and we conducted χ^2 tests. Our results suggest strong correlations and evidence of contingency, especially for the governmentwide analyses. For the five-item stepwise regression, the correlation coefficient was 1.00 and χ^2 was 65 ($df = 1, p < .01$). For the 10-item stepwise regression, the correlation coefficient was $r = .82$ and χ^2 was 39 ($df = 1, p < .01$). In general, the survey key driver items were significantly correlated across years, indicating a degree of temporal generalizability. Thus, there is little empirical evidence to support the practice of conducting SKDA each year that a survey is administered. However, we recognize that follow-up research looking at longer time periods or at nonfederal organizations might yield slightly different results.

Issue 7: Multivariate Analyses Are Suspect When Variables Outnumber Observations

So far we have discussed the use of SKDA for data from large groups of employees. Additional problems can arise when the analysis is conducted for smaller groups, such as teams of workers. A recent article on SKDA suggested that one best practice proposed by SKDA users is to conduct the analysis at the lowest levels within an organization (i.e., the "team" level, Alonso & Wang, 2013, p. 93). In our view, organizations that conduct SKDA at the team level will likely have too few data points to obtain meaningful results. In our experience, most work units have at most 10–15 employees reporting to one supervisor. Consider the 71-item federal survey used in our examples; when completing this survey, the 10–15 employees would respond to a total of 71 predictor items and the criterion. The ratio of respondents to variables (i.e., 10–15 to 71) presents five additional psychometric issues for SKDA.

First, traditional parametric statistics (e.g., Pearson's r) cannot be used effectively with fewer than 30 cases; instead, nonparametric statistics (e.g., Spearman's ρ) should be used. A second issue is statistical power. Even if a work unit was sufficiently large (e.g., at least 30 survey respondents) for parametric statistics, the power for detecting a significant correlation is low; one would need 50 survey respondents in order to detect a significant correlation of .48 (the mean observed correlation in our results) with 95% power. An even greater number of cases are required for

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

sufficient power for determining whether one survey driver correlation is significantly different than another. Using the average interitem correlation from the 2008 dataset (.40) and the average correlation with the criterion (.48), to achieve 95% power, totals of (a) 1,824, (b) 200, and (c) 72 survey respondents would be required to detect (a) small ($q = .10$), (b) medium ($q = .30$), and (c) large ($q = .50$) effect sizes for the differences between the two correlation coefficients (using q as described in Cohen, 1992). (We note that nonparametric statistics further exacerbate the issues with low power since they have less power than parametric statistics.)

Third, in small samples the confidence intervals around a point estimate of a correlation are quite large. With 30 cases and a true correlation of .48, the values within the 95% confidence interval would range from .15 to .72. Fourth, regression-based analyses are problematic when there are more variables (i.e., the 71 items) than cases (i.e., the 10–30 survey respondents per work unit). Both least squares multiple regression and RWA require the inversion of the correlation matrix between the item variables (i.e., the matrix algebra equivalent to taking the reciprocal of a number). Consequently, the correlation matrix in this situation (i.e., 71 variables and 10–30 cases) would be a singular matrix (i.e., the matrix algebra equivalent to zero) and could not be inverted (i.e., the matrix algebra equivalent of dividing by zero). Fifth, even if the number of items was reduced (e.g., to 12) at the work unit level, the ratio of group sample size to survey items entered as predictors (n:k) would yield uninterpretable regression coefficients, if any are estimated (Cohen & Cohen, 1983). In fact, a regression analysis conducted on a set of randomly generated data with n cases, one criterion, and $n - 2$ predictors (which is just barely enough to avoid singularity) will yield a nonsignificant $R^2 \approx 1.0$.

Issue 8: Correlation ≠ Causation

The ultimate aim of SKDA is to identify areas (represented by items) in which an organization can focus action-planning efforts to improve overall job satisfaction or employee engagement. Essentially, SKDA identifies items that are driving job satisfaction or employee engagement up or down within an organization. This interpretation is also consistent with the use of the word "driver," which is defined as "something that makes other things progress, develop, or grow stronger" in the *Cambridge Dictionary* (Cambridge University Press, 2016). Although we realize that not all practitioners who use SKDA believe that the driver items reflect underlying causal issues, the word "drive" and the emphasis on addressing driver items in action planning might imply causation to stakeholders.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

No doubt, the vast majority of I-O psychologists are familiar with the tenet "correlation is not causation," yet the practice of conducting a SKDA and assuming that the driver items are causing job satisfaction or employee engagement seems to contradict this tenet. Unfortunately, causation cannot be established using the observational and correlational approaches inherent in SKDA. If a survey vendor decides to use SKDA, despite the issues we identified above, they should always caveat the findings with a disclaimer that the results are correlational not causational. If we as a field want to provide organizational leaders with the causes of job satisfaction or employee engagement, then we need to conduct experiments (with random assignment) and employ the scientific method to establish causation. Such applied research could prove fruitful for I-O psychologists, who could partner with organizations to conduct the experiments or summarize existing experiments using meta-analysis.

Should I-O Psychologists Conduct SKDA?

SKDA is seductive because it provides the illusion of an easy answer to questions about the causes of an organization's overall level of job satisfaction or employee engagement. In our discussion, we identified eight issues concerning the use of SKDA. Many of these issues incorporate fundamental findings in the psychometrics literature. It remains to be seen whether these issues can be addressed to make SKDA a more issue-free practice. We hope that our discussion will raise the awareness of those using SKDA of the issues stemming from its use. Many may decide to halt use of SKDA, as we have done ourselves. Those who continue to use SKDA should consider its serious limitations and accordingly frame their presentations of survey results to organizational leaders.

Over the course of history, new analysis techniques have evolved. Consider the emergence of item response theory and CFA. Ordinarily when new analyses are developed, they undergo rigorous testing (e.g., Monte Carlo simulations, formulaic derivations) before being used. Ideally, published empirical, theoretical, and simulation papers will appear in the peer-reviewed literature before a new analytic tool is put into practice. In conducting our literature searches for this focal article, we found surprisingly few examples of this work. The review article by Scherbaum et al. (2010) stands out as an exception. Despite the relatively small and insufficient research base to support SKDA, I-O psychologists have been putting SKDA into practice. In our opinion, SKDA is an example of practice moving ahead of—and beyond—science. Also, as suggested by an anonymous reviewer, SKDA may tempt organizational leaders to accept quick and easy answers rather than having to dig deeper and engage in discussions with their employees about the underlying issues impacting them.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

What Other Approaches Exist?

Given the issues with SKDA, practitioners likely will need to turn to other approaches for guiding the development of action plans. Several existing approaches have been described in the literature and have enjoyed extensive use. The Office of Personnel Management (Berry, 2012) has prepared a public-domain document that provides concise guidance on interpreting survey results, conducting follow-up analyses (including the use of focus groups and open-ended surveys), and performing action planning. Though geared toward executives, this document has much to offer I-O practitioners. Other approaches include holding strategic meetings with senior leaders (Church & Oliver, 2006; Wiley, 2010) and asking employees to rate the importance of different issues on their job satisfaction (Church & Waclawski, 1998; Society for Human Resource Management, 2012). These approaches can be used in addition to, or as a replacement for, SKDA.

Our Proposed New Approach

As an alternative to SKDA, we propose that by following the steps of the scientific method, which implies the adoption of the belief that surveys are themselves psychometric instruments similar to selection assessments, I-O psychologists can implement a more sound approach to identifying the root causes of job satisfaction or employee engagement.

Step 1: Identify targets for intervention. As a first step once a survey has been conducted, scientist–practitioners could examine the survey results to identify those organizations (or suborganizations) that would benefit most from action planning (i.e., identify the population to research). The target organization's responses on individual items and on scales would be compared with those of other organizations (e.g., using a normative database). After partialling out any factor-related variance and examining the statistical significance and effect sizes of any differences between the target organization and the normative comparison group, practitioners could identify "red flags" or areas in which the target organization scored substantially lower than comparison groups.

Step 2: Develop hypotheses to explain the survey results. Assuming that the target organization has areas in need of improvement, the next step would be to develop hypotheses about reasons for low scores on items or scales. Existing findings from the organizational psychology research literature provide one source for hypothesis generation. A second source of hypotheses, contextual and local information from the organization, can be gained through structured qualitative research techniques such as free-response survey questions, facilitator-led focus groups, verbal protocols/cognitive interviews with survey respondents (Barber & Roehling, 1993; Ericsson & Simon, 1993; Robie, Brown, & Beaty, 2007; Trickett & Trafton,

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

2009), and interviews with organizational members. For example, in gaining information about organizational communications, one might include the following questions:

> 70% of employees indicated that they do not have enough information to their jobs well. What information are you lacking in order to do your job well? Why do you think you are not receiving the information? What do you think management should do to better inform employees?

Ideally, a survey researcher would collect information from employees at all levels of the organization (e.g., nonsupervisory employees, supervisors/managers, executives). Hypotheses can also be generated by benchmarking the organization's policies and practices against those of similar organizations.

Step 3: Design organizational interventions. Once hypotheses have been generated, organizational interventions can be designed. As is the case for laboratory experiments, some of the interventions may not be feasible or practical to implement. For example, in an organization undergoing bankruptcy proceedings it might not be feasible to increase employee monetary awards. Instead, an intervention that involves informing and educating employees on the realities facing the organization might have to be used.

Step 4: Implement the intervention. Once designed, the intervention can be implemented using Cook and Campbell's (1979) work on quasi-experimental design as a guide. From a scientific perspective, to establish causality, employees should be randomly assigned to either receive the intervention or be in a control group. After a sufficient period has elapsed, the survey would be readministered, and the resulting data for the two groups would be compared using a *t* test and a measure of effect size. If the hypothesized improvement in survey results fails to emerge, the scientist–practitioner must look to an alternate explanation for poor survey results and develop new hypotheses.

Cook and Campbell (1979) aptly describe the difficulties of conducting true experiments in field settings. Their guidance suggests the need to extend an organizational intervention to all employees. However, they warn that improvements in the outcome variable (job satisfaction or engagement results in our case) cannot be attributed to the intervention. The client organization would consider this to be a moot point, if the original problem (i.e., low employee job satisfaction or engagement results) no longer exists. Likewise, if employee job satisfaction or engagement remains low, new hypotheses must be generated. However, it would be impossible to rule out the possibility that the intervention was successful but its effects were offset by another variable.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Toward a theory of the causes of job satisfaction or employee engagement. We recognize that our scientific approach is more laborious than conducting a SKDA and could result in organizational spend levels that exceed feasibility. Our approach is perhaps best considered to be an ideal state. At a minimum, an improvement on the current SKDA methodology that is rooted in scientific study and the use of at least some of the steps serves to enhance the potential for meaningful outcomes and interventions. This scientific approach is consistent with methodology used in other applied settings, such as the use of randomized control trials in the pharmaceutical industry. Further, if I-O psychology as a field can identify interventions that have a causal effect on job satisfaction or employee engagement, there will be less of a need for further experimental work on this topic (i.e., the additional work to identify causes could be a temporary effort). Ideally, the field should aim to develop scientific theories that describe the interventions that cause changes in job satisfaction or employee engagement. This needs to be accomplished via primary studies, replications, and finally meta-analysis. Without experimental work establishing true scientific theories of the causes of job satisfaction or employee engagement, we as a field cannot provide organizations with effective and empirically based guidance.

Conclusion

In summary, we outlined eight issues concerning the use of SKDA that appear neglected in the literature and in practice. Most notably, we established that the identification of an item as a key driver largely depends on the *SD* of responses to that item and its loading on the factors underlying the survey responses. Furthermore, the results of SKDA tended to be stable across different organizations and different survey administrations, suggesting that organizations should not conduct SKDA driver items with each administration. Perhaps most problematic, randomly chosen items predicted the criterion about as well as the driver items! We suggest that practitioners think deeply about conducting SKDA and that the field move to establishing causational inferences about the "drivers" of job satisfaction or employee engagement rather using correlational designs and analyses.

Of course, our article is not intended to be the final word on SKDA. Instead we view it as the beginning of a discussion about the appropriateness of SKDA and more overarching issues concerning the use of organizational survey results to improve organizational functioning. There are a number of questions and topics that I-O psychologists can consider and debate on concerning SKDA, including the following:

- What should I-O practitioners conduct instead of SKDA to answer client's questions about where to focus action-planning efforts?

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

- Why have I-O academics not studied SKDA extensively?
- What scientific evidence exists that SKDA is effective?
- Why are we using statistical analyses when other more direct approaches are available?
- As a common practice, why is the development of organizational surveys not guided by psychometric theory more often?

In closing, we hope that our article will increase awareness of the issues concerning SKDA and prompt the field to think more deeply about the analysis and its alternatives.

References

Alonso, A., & Wang, M. (2013). International practice forum. *The Industrial–Organizational Psychologist, 50*(3), 90–94.

Barber, A. E., & Roehling, M. V. (1993). Job postings and the decision to interview: A verbal protocol analysis. *Journal of Applied Psychology, 78*(5), 845–856.

Berry, J. (2012). *Guide for interpreting and action on Federal Employee Viewpoint Survey results*. Washington, DC: U.S. Office of Personnel Management. Retrieved from http://www.chcoc.gov/transmittals/TransmittalDetails.aspx?TransmittalID=5175

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.

Cambridge University Press. (2016). *Cambridge dictionary*. Retrieved from http://dictionary.cambridge.org/us/dictionary/english/driver

Church, A. H., & Oliver, D. H. (2006). The importance of taking action, not just sharing survey feedback. In A. I. Kraut (Ed.), *Getting action from organizational surveys* (pp. 102–130). San Francisco, CA: Jossey-Bass.

Church, A. H., & Waclawski, J. (1998). *Designing and using organizational surveys: A seven-step process*. San Francisco, CA: Jossey-Bass.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colquitt, A. L., & Macey, W. H. (2005, April). *Surveys throughout the employment lifecycle: What matters, when*. Preconference workshop presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Cook, T. D., & Campbell, D. T. (1979). *Quasi experimentation: Design and analytical issues for field settings*. Chicago, IL: Rand McNally.

Cucina, J. M., Credé, M., Curtin, P. J., Walmsley, P. T., & Martin, N. R. (2014, May). Large sample evidence of a general factor in employee surveys. Poster presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement, 10*, 94–96.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Gast, I. F. (2008, November). *Developing and administering web-based surveys: A tool-kit for assessment professionals*. New York, NY: Mid-Atlantic Personnel Assessment Conference.

Gast, I. F. (2010, July). *Questions you should answer before conducting a survey*. Newport Beach, CA: International Personnel Assessment Council Conference.

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

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*, 238–257.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Kraut, A. I. (2006). Moving the needle: Getting action after a survey. In A. I. Kraut (Ed.), *Getting action from organizational surveys* (pp. 1–32). San Francisco, CA: Jossey-Bass.
Leonard, D., & Lieberman, M. (2006). Data use: Using key driver analysis to guide employee satisfaction research. *Quirk's Marketing Research Review, 20*(9), 20–24.
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* (pp. 326–351). San Francisco, CA: Jossey-Bass.
Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(1), 3–30.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw Hill.
Rasch, R. L., Kowske, B., & Herman, A. E. (2009, April). *Identifying key drivers and actionable items: Comparison of four methods*. Paper presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology, 21*(4), 489–509.
Rodriguez, D., Usala, P., & Shoun, S. (1995). *Occupational analysis of federal clerical and technical occupations: An application of the multipurpose occupational systems analysis inventory—Closed-ended* (MOSAIC; Report No. PRDC-95-08). Washington, DC: U.S. Office of Personnel Management, Personnel Resources and Development Center.
Scherbaum, C. A., Putka, D. J., Naidoo, L. J., & Youssefnia, D. (2010). Key driver analyses: Current trends, problems, and alternative approaches. In S. L. Albrecht (Ed.), *Handbook of employee engagement: Perspectives, issues, research, and practice* (pp. 182–196). Northampton, MA: Edward Elgar.
Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.
Society for Human Resource Management. (2012). *2012 employee job satisfaction and engagement: How employees are dealing with uncertainty*. Alexandria, VA: Author.
Trickett, S. B., & Trafton, J. G. (2009). A primer on verbal protocol analysis. In D. Schmorrow, J. Cohn, & D. Nicholson (Eds.), *The PSI handbook of virtual environments for training and education* (pp. 332–346). Westport, CT: Praeger.
U.S. Department of Labor. (1974). *Descriptive rating scale* (Form MA-7-66, Rev. 3-74). Washington, DC: U.S. Department of Labor, Manpower Administration.
Wiley, J. W. (1996). Linking survey results to customer satisfaction and business performance. In A. I. Kraut (Ed.), *Organizational surveys: Tools for assessment and change* (pp. 330–359). San Francisco, CA: Jossey-Bass.
Wiley, J. W. (2010). *Strategic employee surveys*. San Francisco, CA: Jossey-Bass.
Wiley, J. W., & Campbell, B. H. (2006). Using linkage research to drive high performance: A case study in organizational development. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications* (pp. 150–182). San Francisco, CA: Jossey-Bass.
Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58–79.

Appendix A: Methodology

We used an archival dataset of 212,223 civilian, full-time, permanent, U.S. federal government employees who participated in a federal governmentwide survey in 2008 for most of the analyses. Seventy-three federal government organizations participated in the survey. The dataset was further broken down into 272 suborganizations. The organizations represented have diverse workforces and missions, ranging from law enforcement agencies to scientific organizations and to quasi-public organizations such as the Pension Benefit Guaranty Corporation. One analysis also used datasets for the 2004, 2006, and 2010 administration of the survey. Table A1 provides background information on the four administrations of the survey.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Table A1 Administration Information for Each Survey

Year given	*N*	Number of organizations	Number of sub-organizations	Response rate	Source
2004	147,914	35	70	54%	http://www.fedview.opm.gov/2004/fhcsdata/
2006	221,480	82	299	57%	http://www.fedview.opm.gov/2006/fhcsdata/
2008	212,223	73	272	51%	http://www.fedview.opm.gov/2008/fhcsdata/
2010	263,475	78	279	52%	http://www.fedview.opm.gov/2010/fhcsdata/

The survey items are provided in Appendix B, Table B1. All items used polytomous 5-point Likert scale response formats. Several of the items are used in a consortium of employee survey administrators. The criterion item was "considering everything, how satisfied are you with your organization?" Cucina, Credé, Curtin, Walmsley, and Martin (2014, Study 3) have studied the correlation between responses to the 2012 administration of the survey items and suborganizational level turnover rates. Using their data, the criterion item had a correlation of −.21 ($p < .01$) with turnover rates. Using data from Rodriguez, Usala, and Shoun (1995), Cucina et al. (2014, Study 4) studied the correlation between responses to another federal employee survey and supervisory ratings of job performance (using a research-based administration of the Descriptive Rating Scale; U.S. Department of Labor, 1974). A differently worded version of this criterion item ("Considering everything, how would you rate your overall satisfaction in the organization?") had an average correlation of .13 ($p < .01$) with the seven performance-related questions in the scale. Two other possible criterion items, "Considering everything, how satisfied are you with your job?" (which correlated −.28, $p < .01$, with turnover rates and had an average correlation of .18, $p < .01$, with the seven performance-related questions) and "I recommend my organization as a good place to work" (which correlated −.32, $p < .01$, with turnover rates and was not included in the study with performance data) were not included as predictors. Using these two items as the criterion for SKDA did not change the conclusions reported in the focal article. Additionally, an index was created by summing the three criterion items. This index had a coefficient alpha of .89. When aggregated to the suborganization level, it had a coefficient alpha of .96 and correlated −.27 ($p < .01$) with turnover rates.

Survey key driver items were identified using multiple approaches. Many analyses used individual employees as cases. In these analyses, correlation coefficients between each item and the criterion item were obtained separately for (a) each suborganization, (b) each organization, and (c) across the entire federal government. Second, standardized regression coefficients were obtained when all items were entered into regression equations using the criterion item; the R^2 and adjusted R^2 values were recorded. Third, two stepwise regression analyses (one identifying five items, the other 10 items) were conducted. Fourth, three of the authors rationally rated each item with respect to its relationship with the criterion item and whether change can be actionable at the suborganization, organizational, or governmentwide level. The consistency of the ratings was measured using the intraclass correlation (3, k); a value of .611 (95% CI: .517 to .689) was obtained. Fifth, partial, semipartial, and RWA coefficients were obtained. According to Johnson and LeBreton (2004), two types of RWA statistics are available (raw and rescaled); because the rescaled weights are a linear transformation of the raw weights, we only report results for the raw weights. In additional analyses, the item responses for employees in an organization or suborganization were aggregated to create organizational and suborganizational level data. In these analyses, each case consisted of the average scores (across all employees) for each organization or suborganization on the items.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Other analyses incorporated results of a CFA of the survey by Cucina et al. (2014, Scheme A). Cucina and his colleagues found that a six factor grouping of the items (My Work Experience, My Work Unit, My Agency, My Supervisor/Team Leader, Leadership, and My Satisfaction) yielded the best fit in the 2010 and 2011 dataset. We applied this model to the 2008 dataset, which included 14 additional items not studied by Cucina et al. Three of these additional items were rationally assigned to factors by the first author, and a set of 11 items covering satisfaction with federal employee benefits were assigned to a seventh factor. An oblique model that allowed all seven factors to intercorrelate and included a path from each factor to the criterion was used. Using the unweighted least squares (ULS) estimation procedure in AMOS (Brown, 2006), we tested this model at the individual, sub-organizational, and organizational levels. Polychoric correlations were used for the individual-level analyses (see Wirth & Edwards, 2007), and Pearson correlations were used for the suborganization and organization-level analyses.

Appendix B

Table B1 Survey Items for 2004, 2006, 2008, and 2010 FHCS Datasets

			FHCS item number			
No.	**Stem**	**Scale**	**2004**	**2006**	**2008**	**2010**
1	The people I work with cooperate to get the job done.	A	1	1	1	20
2	I am given a real opportunity to improve my skills in my organization.	A	2	2	2	1
3	I have enough information to do my job well.	A	3	3	3	2
4	I feel encouraged to come up with new and better ways of doing things.	A	4	4	4	3
5	My work gives me a feeling of personal accomplishment.	A	6	5	5	4
6	I like the kind of work I do.	A	7	6	6	5
7	I have trust and confidence in my supervisor.	A		7	7	51
8	I recommend my organization as a good place to work.	A	8	8	8	40
9	Overall, how good a job do you feel is being done by your immediate supervisor/team leader?	B	9	9	9	52
10	How would you rate the overall quality of work done by your work group?	B	10	10	10	28
11	The workforce has the job-relevant knowledge and skills necessary to accomplish organizational goals.	C	12	11	11	29
12	My supervisor supports my need to balance work and other life issues.	A	13	12	12	42

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Table B1 Continued

			FHCS item number			
No.	**Stem**	**Scale**	**2004**	**2006**	**2008**	**2010**
13	Supervisors/team leaders in my work unit provide employees with the opportunities to demonstrate their leadership skills.	A	14	13	13	43
14	My work unit is able to recruit people with the right skills.	A	15	14	14	21
15	The skill level in my work unit has improved in the past year.	A	16	15	15	27
16	I have sufficient resources (for example, people, materials, budget) to get my job done.	A	17	16	16	9
17	My workload is reasonable.	A	18	17	17	10
18	My talents are used well in the workplace.	A	19	18	18	11
19	I know how my work relates to the agency's goals and priorities.	A	20	19	19	12
20	The work I do is important.	A	21	20	20	13
21	Physical conditions (for example, noise level, temperature, lighting, cleanliness in the workplace) allow employees to perform their jobs well.	A	22	21	21	14
22	Promotions in my work unit are based on merit.	A	23	22	22	22
23	In my work unit, steps are taken to deal with a poor performer who cannot or will not improve.	A	24	23	23	23
24	Employees have a feeling of personal empowerment with respect to work processes.	A	26	24	24	30
25	Employees are rewarded for providing high quality products and services to customers.	A	28	25	25	31
26	Creativity and innovation are rewarded.	A	29	26	26	32
27	Pay raises depend on how well employees perform their jobs.	A		27	27	33
28	Awards in my work unit depend on how well employees perform their jobs.	A	30	28	28	25
29	In my work unit, differences in performance are recognized in a meaningful way.	A	31	29	29	24
30	My performance appraisal is a fair reflection of my performance.	A	33	30	30	15

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Table B1 Continued

No.	Stem	Scale	FHCS item number 2004	2006	2008	2010
31	Discussions with my supervisor/team leader about my performance are worthwhile.	A	34	31	31	44
32	In my most recent performance appraisal, I understood what I had to do to be rated at different performance levels (for example, Fully Successful, Outstanding).	A			32	
33	I am held accountable for achieving results.	A	35	32	33	16
34	Supervisors/team leaders in my work unit are committed to a workforce representative of all segments of society.	A	36	33	34	45
35	Policies and programs promote diversity in the workplace (for example, recruiting minorities and women, training in awareness of diversity issues, mentoring).	A	37	34	35	34
36	Managers/supervisors/team leaders work well with employees of different backgrounds.	A	38	35	36	55
37	I have a high level of respect for my organization's senior leaders.	A	39	36	37	61
38	In my organization, leaders generate high levels of motivation and commitment in the workforce.	A	40	37	38	53
39	My organization's leaders maintain high standards of honesty and integrity.	A	41	38	39	54
40	Managers communicate the goals and priorities of the organization.	A	42	39	40	56
41	Managers review and evaluate the organization's progress toward meeting its goals and objectives.	A	43	40	41	57
42	Employees are protected from health and safety hazards on the job.	A	44	41	42	35
43	My organization has prepared employees for potential security threats.	A	45	42	43	36
44	Complaints, disputes, or grievances are resolved fairly in my work unit.	A	46	43	44	
45	Arbitrary action, personal favoritism, and coercion for partisan political purposes are not tolerated.	A	47	44	45	37

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Table B1 Continued

No.	Stem	FHCS item number				
		Scale	2004	2006	2008	2010
46	Prohibited personnel practices (for example, illegally discriminating for or against any employee/applicant, obstructing a person's right to compete for employment, knowingly violating veterans' preference requirements) are not tolerated.	A	48	45	46	38
47	I can disclose a suspected violation of any law, rule, or regulation without fear of reprisal.	A	49	46	47	17
48	Supervisors/team leaders provide employees with constructive suggestions to improve their job performance.	A	50	47	48	46
49	Supervisors/team leaders in my work unit support employee development.	A	51	48	49	47
50	Employees have electronic access to learning and training programs readily available at their desk.	A	52	49	50	
51	My training needs are assessed.	A	53	50	51	18
52	Managers promote communication among different work units (for example, about projects, goals, needed resources).	A	55	51	52	58
53	Employees in my work unit share job knowledge with each other.	A	56	52	53	26
54	Employees use information technology (for example, intranet, shared networks) to perform work.	A	57	53	54	
55	How satisfied are you with your involvement in decisions that affect your work?	C	59	54	55	63
56	How satisfied are you with the information you receive from management on what's going on in your organization?	C	60	55	56	64
57	How satisfied are you with the recognition you receive for doing a good job?	C	61	56	57	65
58	How satisfied are you with the policies and practices of your senior leaders?	C	62	57	58	66
59	How satisfied are you with your opportunity to get a better job in your organization?	C	63	58	59	67

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press

Table B1 Continued

No.	Stem	Scale	FHCS item number 2004	2006	2008	2010
60	How satisfied are you with the training you receive for your present job?	C	64	59	60	68
61	**Considering everything, how satisfied are you with your job?**	C	65	60	61	69
62	Considering everything, how satisfied are you with your pay?	C	66	61	62	70
63	**Considering everything, how satisfied are you with your organization?**	C	67	62	63	71
64	How satisfied are you with retirement benefits?	C	68	63	64	
65	How satisfied are you with health insurance benefits?	C	69	64	65	
66	How satisfied are you with life insurance benefits?	C	70	65	66	
67	How satisfied are you with long term care insurance benefits?	C	71	66	67	
68	How satisfied are you with the flexible spending account (FSA) program?	C	72	67	68	
69	How satisfied are you with paid vacation time?	C	73	68	69	
70	How satisfied are you with paid leave for illness (for example, personal), including family care situations (for example, childbirth/adoption or eldercare)?	C	74	69	70	
71	How satisfied are you with child care subsidies?	C	75	70	71	
72	How satisfied are you with work/life programs (for example, health and wellness, employee assistance, eldercare, and support groups)?	C	76	71	72	
73	How satisfied are you with telework/telecommuting?	C	77	72	73	73
74	How satisfied are you with alternative work schedules?	C	78	73	74	74

Note. Criterion items are shown in **boldface**. Response Option Scale A: 5 = strongly agree; 4 = agree; 3 = neither agree nor disagree; 2 = disagree; 1 = strongly disagree. Response Option Scale B: 5 = very good; 4 = good; 3 = fair; 2 = poor; 1 = very poor. Response Option Scale C: 5 = very satisfied; 4 = satisfied; 3 = neither satisfied nor dissatisfied; 2 = dissatisfied; 1 = very dissatisfied. Responses labeled missing, do not know, and no basis to judge were treated as missing in all of the analyses in this focal article.

https://doi.org/10.1017/iop.2016.97 Published online by Cambridge University Press