Estimating the economic impact of personnel selection tools on counterproductive work behaviors

Saul Fine[1*]

[1]*Midot Ltd, Israel*

Abstract

Well established methods are available for estimating the monetary value of tools used to predict job performance in personnel selection. However, similar methods for estimating the value of tools used to predict counterproductive work behaviors (CWB) have been less well researched. This article presents two possible approaches for studying the latter issue. These approaches are based primarily on rational estimates of SDy for CWB, and on selection tools' true positive rates. Anecdotal validity evidence is then used to demonstrate each approach. While more comprehensive research is still needed in this area, methods such as these may already begin to facilitate better informed personnel decisions for managing counterproductive behaviors in the workplace.

Keywords: utility, counterproductive work behaviors, integrity, selection
JEL Classification Codes: M12, M51

1. Introduction

Nearly all organizations, large and small, are exposed to the potential damages of counterproductive work behaviors (CWB). CWB refers to a broad range of intentional acts that threaten an organization and/or its members (Robinson and Bennett, 1995). Some CWBs affect the organization (e.g., theft, unexcused absences), while others target its employees (e.g., harassment, fighting). When considered collectively, the great majority of employees around the world are estimated to have been involved in some form of CWB during their careers (Coyne and Bartram, 2000; Slora, 1991), and the organizational damages caused by such behaviors have been valued in the hundreds of billions of dollars annually (Murphy, 1993). Employee thefts, for

---

example, may be responsible for roughly 35% of an organization's inventory shrinkage, an average of 1.4% of its total revenues, and more than $38 billion dollars in annual damages for the retail industry alone (Center for Retail Research, 2010).

As substantial as these estimations may be, the actual damages associated with CWB may be significantly higher, due to their relatively low detection rates. While the base rate for employee thefts, for example, may be as high as 50% (Wimbush and Dalton, 1997), detection rates are often less than 5% (Murphy, 1987). As a result, and in light of the clearly significant costs associated with CWB, it is not surprising that many organizations choose to invest in personnel selection tools designed to mitigate the extent of CWB. This paper presents two initial approaches towards measuring the economic impact of such tools.

2. Personnel selection and CWB

While the antecedents of CWB can be traced to characteristic situational factors, such as lax security controls, corrupt organizational cultures, and stressful or unfair working conditions, employees' dispositional traits have also shown to play an important role in predicting CWB (Sackett and Devore, 2001). In fact, while situational interventions aimed at increasing security controls, for example, can be effective for deterring and detecting incidents of CWB among job incumbents, the assessment of individual traits among job applicants may help to prevent potentially problematic employees from being hired in the first place. Indeed, companies around the world administer psychological assessments in their recruitment and selection processes for this very purpose. And, among the most widely used and well researched of these assessments are integrity tests (Ones, Viswesvaran and Schmidt, 1993).

Traditional integrity tests (i.e., "overt" tests) are designed to measure examinees' attitudes and opinions towards CWB, whereby self-reported rationalizations and justifications towards CWB are associated with an increased risk towards future engagement in such behaviors (Murphy, 1993; Wanek, 1999). The attitudes measured by overt integrity tests are considered to develop primarily through an individual's learned social experiences (Bandura, 1991), but may also be related to more stable dispositional traits, such as conscientiousness, emotional stability, and agreeableness (Berry, Sackett and Wiemann, 2007). Meta-analyses estimate the overall operational validity of overt integrity tests for predicting CWB to range from .38 (uncorrected: .30; Van Iddekinge et al., 2012) to .55 (uncorrected: .39; Ones et al., 1993). Accordingly, overt integrity tests may be considered useful tools for personnel selection in a variety of settings.

3. Utility and CWB

A fundamental aspect of the usefulness of any selection tool for an organization is its economic impact. Specifically, valid assessments allow organizations to reliably predict future job performance among job applicants, according to which more accurate hiring decisions can be made. And, since high job performance can be associated with organizational productivity, tools used to select high-potential employees can have a positive impact on an organization's financial performance – a metric known as *utility* (Cascio and Aguinis, 2011).When considering the adoption of assessments designed to predict job performance, for example, organizations can estimate predicted gains in productivity based on the measured competencies of those hired. So too, when considering assessments designed to predict CWB, organizations should be able to

estimate the predicted overall savings from the prevented damages associated with *not* hiring certain high-risk job applicants.

In the case of the former, while several approaches exist to measure utility (see Boudreau and Ramstad, 2002), utility is traditionally calculated as a function of: (a) the test's validity, (b) the standard deviation of monetary-valued performance (SD_y), (c) the selection ratio, and (d) the cost of the test. These parameters are the basis of the classic Brodgen-Cronbach-Gleser (B-C-G) utility formula (and its later adaptations), which have been consistently reported to yield substantial monetary returns, not uncommonly reaching sums in the millions of dollars (Gatewood and Feild, 1998).

Where assessment tools are designed to predict CWB, however, estimating utility using the B-C-G model is less straightforward, for several possible reasons: (a) the B-C-G formula, in its current format, considers only the performance of those hired, and not of those rejected; (b) CWB is considered to be a unique type of job performance, rather than simply a low level of task performance (Sackett and Devore, 2001); and (c) due to the low detection rates of CWB, the extent of individual-level damages may be difficult to quantify in terms of their monetary values.

A pivotal challenge common to all of the above issues seems to be estimating the SD_y of CWB. Traditionally, a variety of methods are available for estimating the SD_y of job performance. Perhaps the most widely referenced methods are based on the global estimations offered by Schmidt et al. (1979), whereby the monetary value of employee performance output is estimated at different percentile levels around the median (i.e., 15th, 50th and 85th percentiles), and the subsequent method by Schmidt and Hunter (1983), according to which SD_y is computed in reference to a job's mean annual salary (typically falling between 40% and 70%). Unfortunately, similar estimates are not readily available for estimating the SD_y of CWB.

This is not to say that the economic value of CWB has not been well studied (see Tziner and Birati, 2002). Rather, these values have generally been studied in terms of their net damages, independent of selection tool validities. In the absence of individual-level models, alternative group-level methods have been used to analyze the utility of selection tools for prevented CWB. Group-level utility estimates include measuring organizational metrics (e.g., behavioral incidents, dismissal rates, inventory shrinkage), before and after implementation of the selection tool (i.e., "time series" analyses), or between samples of individuals hired with or without the selection tool (i.e., "contrasted groups" analyses). The integrity test literature, for example, includes several studies of this kind, showing significant monetary savings following the use of integrity testing in personnel selection (see Jones, 1991; Miner and Capps, 1996). In one such study, for example, Brown et al. (1987) reported a 50% reduction in the number of employee terminations due to theft, illegal drug use, and violence, as well as savings in inventory shrinkage losses of more than two million dollars.

An advantage of group-level analyses is that they are able to directly measure bottom-line cost metrics, and, as such, do not require estimates of SD_y. On the other hand, this may be a potential limitation, as some metrics (e.g., inventory shrinkage, turnover) are not solely the result of CWB. Moreover, changes in these metrics cannot be directly attributed to the selection tool used, as opposed to other confounding factors. Nonetheless, group-level studies are clearly informative, as it would be hard not to attribute at least part of the differences found in these metrics to the selection tools, especially after repeated studies. However, even conceding that group-level studies may indicate utility *ex post facto*, they still do not lend themselves to predicting future utility in new settings among job applicants, prior to the selection tool's implementation.

4. Estimating the utility of prevented CWB

The following sections briefly present two possible methods for estimating the value of prevented losses, potentially saved by rejecting high-risk applicants using valid selection tools. The first method will be referred to here as the "*SD_y method*" and the second will be referred to as the "*true positive method*." The *SD_y* method is based on estimates of the standard deviation of CWB, and the true positive method is based on the mean cost per CWB. In both cases, it should be noted that for sake of simplification, the value of possible indirect damages caused by CWB, such as lowered morale, reduced cooperation, litigation and compensation fees, etc., will not be addressed here.

The SD_y method. This method is an adaptation of the B-C-G formula, with an important difference in that it focuses on rejected, rather than hired applicants. The formula itself, therefore, is very similar to the original, and can be expressed as follows:

$$Utility\ (SD_y\ method) = \left(N * T * r_{xy} * SD_y * Z_x\right) - (N * C)/RR \qquad (1)$$

where: N = the number of *rejected* applicants; T = the average tenure if the *rejected* applicants were otherwise hired; r_{xy} = the validity of the selection tool used to predict CWB (e.g., integrity tests); SD_y = the standard deviation of the monetary damages associated with CWB; Z_x = the average standard test score of the *rejected* applicant group; C = the administration costs per applicant; RR = the rejection ratio (the number of *rejected* applicants to the number of total applicants).

The following case study is offered to demonstrate utility using this model. In a study of 2,456 Israeli job applicants from 13 large-size organizations operating in eight different industries (e.g., finance, retail, manufacturing, staffing, technology), Fine (2010) found a corrected overall validity coefficient of .32 (.26 uncorrected) for an overt integrity test that was measured against self-reported incidents of CWB. Subsequent to this study, subject matter experts (SMEs) from roughly half of the participating organizations ($k = 6$) were interviewed, in order to calculate a very initial estimate of the SD_y of overall CWB. A method based on Schmidt et al.´s (1979) and Schmidt and Hunter's (1983) techniques was adopted, whereby an average employee's salary was used as a reference point from which the SMEs could estimate the overall damages caused by employees who behave worse than 85% of the employees in the company in terms of CWB (i.e., described collectively as coming late or skipping work without permission, taking home merchandise without permission, damaging property, not adhering to company policy, etc.), compared to the damages caused by an average employee (50th percentile). The mean difference between these two ratings was used as a rough estimate of +1 SD_y. The results of this exercise found SD_y to range between 15% and 75% of an employee's annual salary, with a mean of 45% ($SD = 24.3$).

Adding some approximate values to the SD_y method's formula from Fine's (2010) dataset: $N = 313$; $T = 1$ (a default value), r_{xy} = -.32, SD_y = \$11,250 (i.e., 45% of an average salary of \$25,000 year, according to the Israeli Central Bureau of Statistics, 2009, from the time of the study); Z_x = -1.14; C = \$20.00; RR = 0.1274. This yields: (313 * -0.32 * 11,250 * -1.14) – (313 * 20)/0.1274= \$1,235,415, or as much as 26 times the invested costs over the three month period of the study.

Annualized, this utility is estimated to be approximately $4,941,660 for all of the companies participating in the study, or an average of more than $380,000 for each.

Despite its rational appeal and similarity to classical utility models, the primary weakness of this approach is arguably its assumption of normality. A recent study by O'Boyle and Aguinis (2012) demonstrates that many performance criteria in general, and CWB in particular, may follow Paretian (i.e., power law), rather than Gaussian (i.e., normal) distributions. Accordingly, the present estimate of SD_y is likely to misrepresent the prevented damages associated with rejecting low-scoring job applicants.

The true positive method. The second method proposed here is a simplification of the previous formula. Instead of the three main variables therein (i.e., Z_x, SD_y, r_{xy}), this method: (a) treats the test score as dichotomous (i.e., above/below the test's operational cut-score); (b) estimates the monetary value of CWB as the mean loss due to a specific CWB; and (c) uses the test's true positive rate for predicting CWB. This method can be expressed using the following formula:

$$\textit{Utility (true positive method)} = \left(N * \overline{\#}_y * \overline{\$}_y * \%TP_x \right) - (N * C)/RR \qquad (2)$$

where: N = number of *rejected* applicants; $\overline{\#}_y$ = the mean number of CWB incidents committed; $\overline{\$}_y$ = the mean monetary loss incurred from a CWB; $\%TP_x$ = the percentage of the test's true positives for predicting CWB; C = the administration costs per applicant; RR = the rejection ratio.

To demonstrate this formula, a mean value for global retail employee theft incidents ($1,944; Center for Retail Research, 2010) can be used to roughly estimate $\overline{\$}_y$ for a subset of Fine's (2010) sample. Specifically, among the study's retail (and retail manufacturing) industry's applicants ($N = 566$), only 3.0% ($N = 17$) admitted to having stolen from their past employers, for an average of 1.2 times each. Of these individuals, 47.1% ($N = 8$) also had low (below cut-score) integrity test scores. Since a total of 69 individuals had low test scores, the test's true positive rate was 11.6%, and its overall utility can be estimated as: (69 * 1.2 * 1,944 * .116) – (69 * 20)/0.1219 = $7,351. In other words, an annualized utility of $29,404, or a 165% return on the organizations' investment costs (for thefts alone).

A clear advantage of the true positive model is its simplicity, whereby complex estimations of SD_y are not required. As such, this method's utility estimates may possibly be perceived by managers as more realistic than typical SD_y -based estimates, which are sometimes viewed incredulously, and only marginally (or even negatively) influence human resource decisions (Latham and Whyte, 1994). In addition, the true positive method is not subject to assumptions of normality, which may be questionable with regards to CWB distributions. Another advantage of this method is that it lends itself to aggregating the utility gained by several specific types of CWBs. Were it possible to have estimated the monetary value of additional CWBs via SME judgments, for example, a more global estimate of the test's utility could have been produced. Unlike the SD_y method, however, the disadvantages of this model are that it does not take into consideration the variance (i.e., the individual differences) within the estimates of CWB or the test scores, nor does it consider the true relationship between these two variables (which is naturally continuous, not dichotomous).

5. Concluding Remarks

This paper briefly puts forth two alternative approaches for estimating the monetary value of prevented CWB in personnel selection, which were derived from classic utility models in I/O psychology. Far more advanced adaptations of the original B-C-G model are available, which incorporate additional issues, such as taxes and interest rates (Boudreau, 1983; Tziner and Birati, 2002). These issues were not discussed here for the sake of brevity, although adding them to the above models may substantially influence the utility estimates yielded. In some cases, for example, the losses prevented by a given selection tool may be offset by possible tax deductions allowed for such losses, as in the case of inventory shrinkage. In other cases, interest may be earned on the savings that would have otherwise been needed to cover the losses. Finally, in cases where incidents of CWB would have resulted in employee dismissals, the utility formulae should consider incorporating turnover costs as well. Turnover costs are widely considered to be equivalent to at least 50% of an employee's starting annual salary (Johnson, Griffeth and Griffin, 2000). And, when added to the true positive method's example above, the $N = 8$ true positives identified during the 3-month study (i.e., 32 per year), each earning an average of salary of $25,000, would equal approximately $400,000 in saved turnover costs per year – a sum far superseding the utility associated with prevented thefts alone.

Another aspect not dealt with directly above, but which may also significantly affect utility estimates, is tenure. In traditional utility models, where the mean tenure of successful employees is greater than one year, the test's utility is multiplied as a result. In the case of CWB, however, serious incidents (e.g., thefts) may result in early dismissals, and thereby fraction the estimated utility accordingly. In the SD_y method's example above, tenure data were unfortunately not available, and tenure was therefore set to one as a means to relate to utility on an annual scale. However, if the average tenure of employees involved in CWB was actually 6 months, for example, the utility estimates would be cut in half.

In terms of using the two models presented herein operationally, each has its advantages and disadvantages (as briefly described above), and might therefore be adopted as the relevant data are available. By nature, the SD_y method still requires more comprehensive rational estimates from additional SME samples, before its estimates could be generalized across settings. Moreover, in doing so, alternative techniques for estimating SD_y should be explored. Indeed, several of the SMEs found it very difficult to accurately estimate global SD_y values, raising possible concerns for the reliability of such a measure. Towards a possible solution, future research may attempt to breakdown SD_y estimates into specific behaviors, in a similar approach to that proposed by Cascio and Ramos (1986). Furthermore, a direct investigation of the effects that Paretian distributions of CWB may have on Gaussian-based utility estimates is certainly in order. Assuming these effects are indeed significant, it may be necessary to more accurately estimate SD_y in terms of nonlinear values across multiple test score indices, rather than as single overall estimates. Once established from these perspectives, the SD_y model has the potential to offer more realistic estimates of utility than the true positive method. In the meantime, however, the true positive method may be a more straightforward and useable approach.

A final yet important limitation of the above approaches for estimating utility is related to one of the central criticisms of selection tools designed to predict CWB (e.g., integrity tests) –high false positive rates. Specifically, due in part to the low detection rates of CWB, it is not uncommon for integrity tests' false positive rates to exceed 80%, while still being able to maintain significant validity coefficients (and true positives) overall (Karren and Zacharias, 2007). Hypothetically, a test that rejects 20% of its examinees for risk of future theft, where the

base rate for detected thefts is just 5%, for example, may correctly predict 100% of the future theft offenders. However, 75% of those rejected will be false positives. Similarly, from the previous theft example based on Fine (2010), despite the test's significant validity for thefts alone ($r = 0.21$), and the test's relatively low cut-score (12.2%), only 3.0% of the sample admitted to stealing from their employers. As a result, whereas nearly half of the admitted offenders were correctly identified by the test as low scorers, the great majority of all low scorers (i.e., 88.4%, or 10.8% of the total sample) were false positives. In response to this issue, one could argue the statistics to be exaggerated, due to the fact that many other types of CWB, other than just theft, need to be considered simultaneously. One could also argue that not all theft offenders were reliably measured by the criterion (i.e., unadmitted thefts). Lastly, one could argue that false positives are part-and-parcel of any selection process, and that in the absence of more valid alternatives, tests of this kind will always facilitate more accurate decisions than the alternative of not using them at all (Sackett and Wanek, 1996). Irrespective of the above arguments, however, in terms of utility, researchers and practitioners may wish to consider the potentially lost *gains* caused by falsely rejecting otherwise productive workers (i.e., the false positives). It is essentially possible, for example, that in at least some cases, the lost profits caused by not hiring false positives may even outweigh the damages caused by not hiring true positives.

References

Bandura, A. (1991) Social cognitive theory of moral thought and action, in W.M. Kurtines and J.L. Gewirtz (eds.) *Handbook of moral behavior and development,* 1, 45-103, Hillsdale, NJ: Erlbaum.

Berry, C.M., Sackett, P.R. and Wiemann, S. (2007) A review of recent developments in integrity test research, *Personnel Psychology,* 60, 271-301.

Boudreau, J.W. (1983) Economic considerations in estimating the utility of human resource productivity improvement programs, *Personnel Psychology*, 3, 551-576.

Boudreau, J.W. and Ramstad, P.M. (2002) Strategic I/O psychology and the role of utility analysis models, CAHRS Working Paper #02-16, Ithaca, NY: Cornell University, ILR School.

Brown, T.S., Jones, J.W., Terris, W. and Steffy, B.D. (1987) The impact of pre-employment integrity testing on employee turnover and inventory shrinkage losses, *Journal of Business and Psychology,* 2, 136-149.

Cascio, W.F. and Aguinis, H. (2011) Applied psychology in human resource management, 7[th]ed., Upper Saddle River, NJ: Pearson Prentice Hall.

Cascio, W.F. and Ramos, R.A. (1986) Development and application of a new method for assessing job performance in behavioral/economic terms, *Journal of Applied Psychology,* 71, 20-28.

Center for Retail Research (2010) The global retail theft barometer 2010, Nottingham, UK.

Coyne, I. and Bartram, D. (2000) Personnel managers' perceptions of dishonesty in the workplace, *Human Resources Management Journal,* 10, 38-45.

Fine, S. (2010) Pre-employment integrity testing across multiple job industries, *Psychological Reports,* 107(2), 1-4.

Gatewood, R.D. and Feild, H.S. (1998) Human resource selection, 4^th^ed.*,* Fort Worth, TX: Harcourt Brace.

Israel Central Bureau of Statistics (2009) Average wages per employee job of Israeli workers, July-September 2009, retrieved July 10, 2012 from http://www.cbs.gov.il.

Johnson, J., Griffeth, R.W. and Griffin, M. (2000) Factors discriminating functional and dysfunctional sales-force turnover, *Journal of Business and Industrial Marketing,* 15(6), 399-415.

Karren, R.J. and Zacharias, L. (2007) Integrity tests: critical issues, *Human Resource Management Review,* 17, 221-234.

Latham, G.P. and Whyte, G. (1994) The futility of utility analysis, *Personnel Psychology,* 47, 31-46.

Miner, J.B. and Capps, M.H. (1996) How honesty testing works*,* Westport, CT: Quorum.

Murphy, K.R. (1987) Detecting infrequent deception, *Journal of Applied Psychology,* 72, 611-614.

Murphy, K.R. (1993) Honesty in the workplace, Pacific Grove, CA: Brooks/Cole Publishing.

O'Boyle, E. and Aguinis, H. (2012) The best and the rest: revisiting the norm of normality of individual performance, *Personnel Psychology,* 65, 79-119.

Ones, D.S., Viswesvaran, C. and Schmidt, F.L. (1993) Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance, *Journal of Applied Psychology,* 78, 679-703.

Robinson, S.L. and Bennett, R.J. (1995) A typology of deviant workplace behaviors: a multidimensional scaling study, *Academy of Management Journal,* 38, 555-572.

Sackett, P.R. and Devore, C.J. (2001) Counterproductive behaviors at work, in N. Anderson, D. Ones, H. Sinangil and C. Viswesvaran (eds.) *Handbook of industrial, work, and organizational psychology,* 1, 145-164.

Sackett, P.R. and Wanek, J.E. (1996) New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection, *Personnel Psychology,* 49, 787-829.

Schmidt, F.L. and Hunter, J.E. (1983) Individual differences in productivity: an empirical test of estimates derived from studies of selection procedure utility, *Journal of Applied Psychology,* 68, 407-415.

Schmidt, F.L., Hunter, J.E., McKenzie, R. C. and Muldrow, T.W. (1979) Impact of valid selection procedures on work-force productivity, *Journal of Applied Psychology,* 64(6), 609-626.

Slora, K.B. (1991) An empirical approach to determining employee deviance base rates, in J.W. Jones (ed.) *Preemployment integrity testing: current research and future directions,* 21–38, NewYork: Quorum Books.

Tziner, A. and Birati, A. (2002) Assessing the financial value of worker organizational behaviors and human resource management programs and interventions, in A. Tziner (ed.) *Human resource management and organization behavior,* 321-367, Aldershot, UK: Ashgate.

Van Iddekinge, C.H., Roth, P.L., Raymark, P.H. and Odle-Dusseau, H.N. (2012) The criterion-related validity of integrity tests: An updated meta-analysis, *Journal of Applied Psychology*, 97(3), 499-530.

Wanek, J. E. (1999) Integrity and honesty testing: what do we know? How do we use it?, *International Journal of Selection and Assessment,* 7(4), 183-195.

Wimbush, J.C. and Dalton, D.R. (1997) Base rate for employee theft: convergence of multiple methods, *Journal of Applied Psychology,* 82, 756-763.