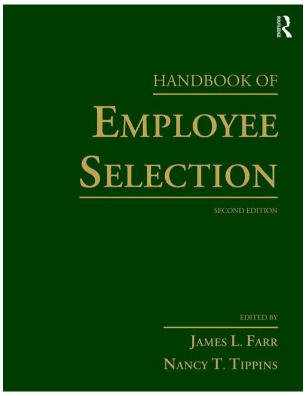
This article was downloaded by: *10.2.97.136*
On: 26 Sep 2023
Access details: *subscription number*
Publisher: *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Employee Selection

James L. Farr, Nancy T. Tippins, Walter C. Borman, David Chan, Michael D. Coovert, Rick Jacobs, P. Richard Jeanneret, Jerard F. Kehoe, Filip Lievens, S. Morton McPhail, Kevin R. Murphy, Robert E. Ployhart, Elaine D. Pulakos, Douglas H. Reynolds, Ann Marie Ryan, Neal Schmitt, Benjamin Schneider

The Sum of the Parts

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-17
Juliet R. Aiken, Paul J. Hanges
Published online on: 22 Mar 2017

How to cite :- Juliet R. Aiken, Paul J. Hanges. 22 Mar 2017, *The Sum of the Parts from:* Handbook of Employee Selection Routledge
Accessed on: 26 Sep 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-17

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://test.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

17

The Sum of the Parts

Methods of Combining Assessments for Employment Decisions

JULIET R. AIKEN AND PAUL J. HANGES

Organizations commonly use multiple assessments (e.g., a combination of cognitive tests, personality tests, situational judgment tests, interviews, etc.) to make hiring decisions. While using more than one assessment to make employment decisions can provide organizations with a more holistic view of each candidate, deciding on how to combine these assessments can have profound consequences on who gets hired. Most employers use one of three approaches to combining assessments: (1) combining multiple assessments subjectively (clinical assessment), (2) developing assessment composites empirically, or (3) using assessments in sequence—i.e., a multiple-hurdle approach. In this chapter, we will define each of these techniques, review when it tends to be used, discuss its strengths and weaknesses, and provide recommendations for its use. In addition to discussing these three methods, we will also discuss how these methods can be used in combination. But, let us begin at the beginning—we will next review each of the three methods of score combination, beginning with clinical assessment.

CLINICAL ASSESSMENT

In clinical assessment, an employer subjectively combines multiple assessments to develop an overall impression of the candidate. Clinical assessments are often based on intuition. Their use may be prompted by the assumption that quantitative approaches to combining assessments fail to capture the complexity of each applicant's qualifications and potential. Despite repeated empirical evidence that clinical assessment tends to have fairly low validity in a number of contexts (Morris, Daisley, Wheeler, & Boyer, 2014), especially compared to statistical decision-making approaches (Ægisdóttir et al., 2006, Grove, Zald, Lebow, Snitz, & Nelson, 2000; Kuncel, Klieger, Connelly, & Ones, 2013), it remains a popular technique for selection and assessment in many workplaces (Highhouse, 2008).

Employers use clinical assessment in a variety of business activities, including leadership or management coaching, individual development assessment, and forming subjective summary scores from applicants' interviews or assessment centers. Clinical assessment is especially popular for selecting candidates into executive positions (Thornton, Hollenbeck, & Johnson, 2010), as job requirements for these positions are seen as specific to that particular opening or organization (Hollenbeck, 2009).

The goal of clinical assessment is to summarize and combine multiple pieces of information about each candidate before making an employment decision; employers using clinical assessment do so to form a holistic judgment of each candidate. There are three phases in clinical assessment: (1) collecting information, (2) evaluating information, and (3) developing a report and recommendation for an individual job candidate (Weiner, 2003). Employers vary on the processes they follow to accomplish each of these steps. For example, information-collecting techniques may lack structure or may be highly standardized. Likewise, information integration varies substantially, as clinical assessment is very subjective. Consequently, who the assessors are and what training they have profoundly influences the assessment process (Morris et al., 2014). Unsurprisingly, the subjectivity of the integration process can make it especially difficult to tell how clinical assessment is conducted in practice.

Clinical assessment is also, due to its subjectivity, both flexible and fairly simple to implement provided that the assessor is not overwhelmed with information on each candidate. However, the drawbacks of clinical assessment far outweigh its strengths. Because clinical assessment is unstandardized, it is typically applied inconsistently both over time and over candidates. Likewise, without any attempt at standardization or objectivity, assessor biases can strongly influence who is hired. Furthermore, clinical assessment particularly lacks validity when assessing candidates for nonmanagerial positions (Morris et al., 2014). Finally, it can be challenging to form a clinical assessment when the assessor has too many different pieces of information to consider at once.

In general, we—like many other scholars—do not recommend that organizations use clinical assessment for selection purposes, even in low-frequency hiring contexts, such as when dealing with small candidate pools, circumstances where few offers are extended, or one-time hiring decisions. While clinical assessment provides useful individualized feedback and can help identify areas for individual growth, it is too subjective for organizations using it to consistently hire or promote the most qualified applicants. Despite this word of caution, if an organization chooses to use clinical assessment, we have a number of recommendations to improve its usage. Specifically, we suggest that organizations (a) use a cognitive ability test in their clinical assessment, (b) standardize the evaluation process, including documenting strengths and weaknesses for each applicant, (c) take structured notes during the evaluation, (d) provide training for evaluators to reduce personal biases, (e) use the same rater or panel of raters to assess all applicants, and (f) only use clinical assessment when determining whom to hire for a high-level managerial position (e.g., executive management).

First, if a clinical assessment approach is taken, organizations should use a cognitive ability test as part of their battery. In their meta-analysis on clinical assessment, Morris et al. (2014) found that clinical assessments were more valid when they included a cognitive ability test. Second, it is critical that organizations standardize their evaluation process as much as possible. This may at first seem counterintuitive; after all, isn't the point of a clinical assessment to make gut decisions, rather than decisions guided by a set rule? However, some structure and standardization can enable decision makers relying on clinical assessment to fight their unseen biases. Our primary recommendation in standardizing the process is to make written reports of each candidate's strengths and weaknesses. These reports should require the evaluator to justify the conclusions they reach, thus helping evaluators think more deeply and thoroughly about their decisions. In addition to creating reports on each applicant, we recommend that employers using clinical assessment also take structured notes during the evaluation or interview to ensure that the evaluator is considering all aspects of the person's strengths and weaknesses in real time. These notes can be used as the framework of the reports.

Furthermore, clinical assessments have higher validity when used to hire candidates into managerial positions (Morris et al., 2014), and particularly into executive positions (Thornton et al., 2010). Thus, if an employer must use clinical assessment, we recommend that it is only used for selecting into high-level managerial roles. Finally, the use of multiple evaluators should be carefully considered when conducting clinical assessment. Specifically, using multiple assessors does *not* improve validity, except when the same assessors are used across all candidates (Morris et al., 2014). Therefore, if multiple evaluators are going to assess candidates, we recommend

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

that organizations use the same evaluators to assess all candidates. We further recommend that organizations employing a clinical assessment approach train their evaluators for consistency. In particular, we recommend frame-of-reference training (Bernardin & Buckey, 1981). Frame-of-reference training involves educating assessors on what attributes or behaviors are desired, providing them with opportunities to practice evaluating, and giving them feedback on their accuracy (Pulakos, 1986). Frame-of-reference training has been shown to reduce the influence of personal biases (Woehr & Huffcutt, 1994) and increase rater consistency (Schleicher, Day, Mayes, & Riggio, 2002).

In summary, although the subjectivity of clinical assessment makes it a less valid method for combining multiple assessments when making employment decisions, employers continue to use it because it is intuitive, because they are concerned it would not be practical to develop an elaborate process for positions with few or rare hires, or because they are skeptical of the ability of hard data to truly identify who will be a good fit for their organization's needs. Two other methods for candidate selection do not suffer from these setbacks, or at least not to the same degree. We turn next to one of these methods. Specifically, the next procedure for weighing multiple pieces of information we will discuss is a compensatory approach.

COMPENSATORY METHODS

Compensatory methods of weighing multiple assessment criteria involve mathematically weighting each piece of information about a candidate (e.g., each assessment score) to determine an overall qualification score for that candidate. These methods are considered "compensatory" because low scores on one assessment can be counterbalanced by high scores on another assessment. Methods of weighting the criteria include unit weighting, regression weighting, factor analysis (e.g., Kanfer, Wolf, Kantrowitz, & Ackerman, 2010), and relative importance analysis, among others.

Typically, compensatory methods are used to select employees into jobs for which weaknesses in one area can be compensated by strengths in another area. For example, let's assume two candidates take six assessments that are scored from 1 to 5 each. The first candidate scores a 1, 3, 4, 5, 5. The second candidate scores a 4, 3, 3, 4, 4. Assuming no minimum score was required on any given assessment, which is a typical assumption in compensatory selection procedures, these two candidates would have an equivalent sum score. The first candidate's strengths in the latter two assessments offset his or her weakness in the first assessment. Both would be equally qualified, assuming equal weights were put on each assessment. However, if we know that anyone who scores below a "3" on the first assessment would not be qualified for the job, then only the second candidate would qualify. Thus, when minimum scores are required, compensatory methods may not be ideal. However, compensatory methods would be well suited for selection in a context where there is no minimum required score on any given assessment.

While compensatory methods are widely thought of as more valid methods for combining predictors than clinical assessment, there are several challenges associated with implementing these methods effectively. First is the obvious issue of how to weight different predictors. Multiple options, including regression weighting, rational weighting, unit weighting, and even Pareto-optimal weighting exist. Each of these methods differs not only in how exact weights are calculated but also in the rationale for its use. We will provide a brief overview of some of the most common methods of weighting predictors next.

Regression weighting each assessment to create a composite involves regressing assessment scores onto the criterion (or criterion composite), then using the resultant regression weights to determine how much to weight applicant scores on each assessment. In contrast, unit weighting involves assigning each predictive assessment a "1"; organizations using unit weighting simply average together standardized scores on each assessment to create a composite. Organizations using rational weighting derive weights for each assessment from a job analysis. Assessments are therefore weighted according to the importance job analysis establishes for each. Finally, factor analyses of measures can be used to determine weights for each assessment within different domain composites (e.g., "ability"; Kanfer et al., 2010).

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

Each of these techniques has different strengths and weaknesses. Regression weighting, for example, is limited in that it assumes there is a linear relationship between KSAOs/competencies and job performance. Furthermore, the rationale behind each weighting practice varies significantly. For example, unit weighting and regression weighting are very different approaches. Specifically, unit weighting focuses on content validity—does the composite reflect all of the desired job components? In contrast, regression weighting focuses on criterion-based validity—does the composite accurately capture the most predictive regression equation?

The rationale behind rational weighting also contrasts with the rationale underlying regression weighting. Specifically, weights derived from rational weighting are imbued with the values of the organization and decision makers, and weights derived from this process would at best implicitly account for predictor and criteria intercorrelation (Hough, Oswald, & Ployhart, 2001). In contrast, regression-derived weights account for intercorrelation explicitly, and the composite is optimal in the mathematical sense rather than directly reflecting organizational values (Hough et al., 2001).

Further complicating the question of how to form compensatory composites is the question of whether and how much the predictors are interrelated. Specifically, the weights recommended by different methods of forming composites diverge depending on the extent to which predictors are orthogonal (Hough et al., 2001). That is, since rational weights do not explicitly take intercorrelation into account, they are likely to differ most dramatically from statistically derived weights when predictors are highly correlated. Indeed, when predictors are highly correlated, regular ordinary least squares (OLS) regression may not be ideal for determining weights. Instead, one statistical technique for minimizing the challenges presented by intercorrelated predictors is relative importance analysis or dominance analysis. Either relative importance or dominance analysis would allow organizations to determine statistical weights (rather than rational weights) for each predictor in the context of the selection model while simultaneously minimizing the effects of suppression and multicollinearity on statistically derived weights (Hough et al., 2001).

Another issue that arises when using compensatory methods of combining predictors is what the organization should use as the criterion (Hattrup & Rock, 2002). Specifically, should the criterion also be a weighted composite? Research on this question reveals that using weighted criterion composites rather than a single criterion assessment can also boost validity and help reduce adverse impact. Unit weighting for criteria appears to result in higher synthetic validity coefficients (Johnson & Carter, 2010). Additionally, weighting job components using multiple regression to create a criterion composite and then weighting predictors based on the criterion composite boosts validity while also reducing adverse impact (Hattrup, Rock, & Scalia, 1997).

Notably, the greatest gains in validity occur when the weights used to form a predictor composite correspond to the values placed on each component in the criterion composite (Hattrup & Rock, 2002). Furthermore, adverse impact is reduced if criteria that correlate with cognitive ability (e.g., task performance) are given lower weights than criteria that correlate less with cognitive ability (e.g., contextual performance) in a criterion composite (Hattrup & Rock, 2002). Of course, as in many areas of scientific inquiry, matching predictor and criterion complexity improves the validity of selection outcomes. Specifically, more complex criteria are predicted better by complex predictor composites that match the criteria on relevance and bandwidth (Hough & Ones, 2001).

There is one final downside to the use of regression to determine predictor or criterion weights that warrants attention. Specifically, when regression weights are used to determine the composite, these weights may be sample and time dependent. Counteracting this concern requires organizations to collect more data, which may be easier said than done; the influence of sample fluctuations lessens with very large sample sizes.

As should be obvious from our discussion thus far, organizations need to make several decisions when developing composite predictors. With so many weighting strategies available, how does an organization know which to use? Unfortunately, most research on weighting strategies does not give an unambiguous and consistent answer to this question. While some research suggests that unit weights are appropriate for combining predictors (Bobko, Roth, & Buster, 2007), other research reports gains in validity coefficient when weighting using other criteria, such as the number of job components or relative weights analysis (Johnson & Carter, 2010).

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

Recently, however, De Corte and colleagues have developed a promising method for compensatory selection that enables organizations to make maximally informed decisions based on Pareto-optimal predictor weights (De Corte, Lievens, & Sackett, 2007; De Corte, Lievens, & Sackett, 2008; De Corte, Sackett, & Lievens, 2010). These weights provide Pareto-optimal tradeoffs between validity and adverse impact (De Corte et al., 2007; De Corte et al., 2008; De Corte et al., 2010). Specifically, weights are considered Pareto-optimal when the level of one outcome (e.g., adverse impact) cannot be improved without losing ground on the other outcome (e.g., decision quality).

More than one outcome is desired in many selection decisions. Ideally, organizations want to develop assessments that maximize validity with regard to relevant outcomes (e.g., performance, organizational citizenship behaviors, etc.) while minimizing adverse impact. However, as we have noted, these outcomes are typically in conflict. Consequently, the procedure for determining Pareto-optimal weights does not result in a single recommended set of weights for each predictor. Instead, the procedure provides a range of possible weights that organizations choose among to reach the desired levels and tradeoff between adverse impact and validity. The method De Corte et al. (2007, 2008, 2010) propose allows organizations to answer concretely how much of an improvement they can make in one of these areas within a given constraint (e.g., 1%, 5%) or penalty imposed in the other area. Additionally, this method provides organizations with information on the worst possible tradeoffs and the relative importance of adverse impact as an outcome for each Pareto-optimal solution maximizing validity (De Corte et al., 2007). Finally, each Pareto-optimal solution maximizes a combined adverse impact-decision quality goal.

Finding Pareto-optimal solutions is a multistep process. The solutions produced by this program first seek to maximize one outcome and then the other. For example, you might first specify that you would like a mean standardized performance level of 0.75 among the pool of candidates who are selected. Multiple combinations of predictors and predictor weights would yield this desired level. That is, multiple combinations maximize your first outcome, validity. Then, you would need to consider which of these combinations would maximize your second outcome (i.e., minimize adverse impact). The combination that maximizes your second outcome (e.g., adverse impact) at each specified level of the first maximized outcome (e.g., performance) is Pareto-optimal (De Corte et al., 2007).

To conduct their proposed analyses, organizations need to specify (a) the selection rate, (b) the representation of minority and majority candidates in the applicant pool, (c) the effect size of the available predictors, (d) the validity of the available predictors, and (e) the intercorrelations of the available predictors (De Corte et al., 2007). Ideally, these estimates are readily available from past or current validation studies or meta-analyses (De Corte et al., 2007). Fortunately, however, the proposed procedure is fairly robust to uncertainty when precise estimates are not available (De Corte et al., 2007). In addition to requiring the specified information, this procedure requires the assumption that predictor and criterion scores have a joint multivariable normal distribution with the same variance-covariance and different means in the minority and majority populations (De Corte et al., 2007).

A detailed explanation of how this procedure works is available in De Corte et al. (2007). Researchers and organizations who want to use this procedure can access a Windows-compatible computer program designed to run it, as well as instructions on how to use this program, at http://users.ugent.be/_wdecorte/software.html. Program users will have several control options, including (a) operationalizing the selection quality objective either by the validity of the composite, the average criterion score of selected applicants, or the utility of the selection; (b) determining the number of tradeoff points computed; (c) specifying if the selection decision is probationary; (d) constraining predictor weights; and (e) specifying upper/lower boundaries or fixing the proportion of hired employees (De Corte et al., 2007).

Although some challenges and caveats are associated with creating valid predictor composites, there is one overarching strength to using these methods that has already surfaced in our previous discussion: the reduction of adverse impact. The preponderance of research shows that most predictors have an adverse impact-validity tradeoff (Pyburn, Ployhart, & Kravitz, 2008). That is, predictors that tend to have high predictive validity also often have high adverse impact, whereas predictors that tend to have lower adverse impact also often have lower validity. In other words, no one predictor is perfect. Compensatory methods of selection allow organizations to address the imperfections of any given selection instrument by combining predictors with different strengths.

A particularly popular compensatory technique used to combat the adverse impact-validity tradeoff is combining noncognitive predictors with cognitive predictors (De Soete, Lievens, & Druart, 2012). In doing so, organizations seek to offset the higher amount of adverse impact typically associated with cognitive predictors with the lower amount of adverse impact associated with noncognitive predictors. Thus, compensatory methods are commonly used to combine personality tests with cognitive tests. However, while it is possible for these combinations to reduce subgroup differences when the assessments are uncorrelated, they are unlikely to reduce those differences as significantly as expected and may even exacerbate differences among groups when assessments are moderately correlated (Sackett & Ellingson, 1997; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997).

In summary, organizations often form statistical composites of assessments to obtain an overall estimate of a candidate's qualifications. These approaches are considered compensatory, since high scores on one assessment can balance out lower scores on another assessment. Compensatory selection procedures are often used in an attempt to balance the goals of low adverse impact and high decision quality. While compensatory approaches have a number of strengths, they can also raise challenging questions. For example, organizations need to consider not only how to form the predictor composites (e.g., rational, unit, regression, Pareto-efficient weighting) but also whether and how to form a criterion composite. We suggest that organizations calculate Pareto-efficient weights when determining predictor composites in order to obtain the preferred balance between adverse impact and predicted performance. Next, we discuss a noncompensatory technique: the multiple-hurdle approach to selection.

MULTIPLE HURDLE

Finally, employers who want to use more than one assessment to determine which applicants are most qualified may turn to a multiple-hurdle approach. In multiple-hurdle selection approaches, employers sequence assessments rather than combining assessment scores to determine whom to hire. For example, rather than weighting a cognitive ability test and a personality test together to determine one overall qualification score, multiple-hurdle procedures might require applicants to first pass an IQ test and then pass a personality assessment. Thus, multiple-hurdle approaches are considered noncompensatory. In order to succeed in a multiple-hurdle testing environment, applicants need to have high scores on all assessments. In contrast, as we have discussed previously, in order to succeed in a compensatory testing environment, high scores on one assessment can balance out lower scores on another assessment.

In addition to their noncompensatory nature, multiple-hurdle approaches to selection rely on other underlying assumptions. One particularly salient assumption is that there is a nonlinear model between knowledge, skills, and abilities, and job performance. The compensatory approach is linear—simply weight each assessment and map a line of predicted performance using that weighted assessment. In contrast, the multiple-hurdle approach assumes that assessments cannot be so easily combined to linearly predict performance. A second, related, key assumption—assuming a classic approach to multiple-hurdle selection—is that there are only two groups of applicants: acceptable and not acceptable. In other words, everyone who passes the set cutoff score for each assessment instrument is considered equally desirable.

Multiple-hurdle assessment approaches are used frequently in practice. They are often used when a large number of people apply for a given job. By using a multiple-hurdle approach in these situations, employers are able to save money by applying assessments to an ever-decreasing number of applicants. Multiple-hurdle assessments are also used when highly technical tests are used. Specifically, if an employer is using an assessment center or other highly involved assessment, using multiple-hurdle assessment techniques in an appropriate sequence may save both time and money.

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

There are a number of strengths to the multiple-hurdle approach to selection. One strength is that it is easy to weed out large numbers of applicants early in the process. As mentioned earlier, doing so allows organizations to reserve more expensive predictors for the most promising applicants. Additionally, using a multiple-hurdle approach strategically may help organizations reduce their adverse impact (De Corte, Lievens, & Sackett, 2006; Sackett & Roth, 1996).

While multiple-hurdle approaches to selection are appealing for reducing expenditures and adverse impact, their use also has several drawbacks. One challenge is where and how to set the cutoff. As discussed, applicants above the cutoff will be considered interchangeable. Therefore, setting a cutoff takes the multiple-hurdle approach away from an emphasis on criterion-related validity (in which a linear model is assumed, and higher scores always more qualified). Instead, the multiple-hurdle approach to selection adopts a content validity approach, where having an adequate amount of each knowledge, skill, or ability is more important than having more of each KSAO. Also, since applicants are treated interchangeably past the cutoff, it is possible that employers do not get the top performers if the underlying relationship between the construct measured by the assessment and performance is linear. See Chapter 8 (this volume) for additional discussion of setting cutoff scores within the multiple-hurdles framework.

We have several recommendations for employers who wish to use a multiple-hurdle approach to selection. First, to reduce adverse impact, prevailing wisdom has held that tests with more adverse impact should be administered later in the hurdle process, and tighter selection should occur in the first hurdle (Sackett & Roth, 1996). In other words, popular belief indicates that the first stage of a selection process should employ low-adverse-impact tests, and comparatively few applicants should make it beyond this stage. The second stage on this smaller group should then employ a higher-adverse-impact assessment, and proportionally fewer applicants should be weeded out by this assessment. By being selective in the first hurdle—which had less adverse impact—and then applying the higher-adverse-impact assessments, organizations can reduce adverse impact with limited loss in predictive validity.

However, a simulation suggests that this wisdom may not hold in all scenarios. Specifically, if predictors have roughly the same validity but differ in adverse impact, then high-impact assessments should precede lower-impact assessments and selectivity should be equal or less severe in the first stage (De Corte et al., 2006). That is, when predictors differ in adverse impact but not in validity, adverse impact is reduced by taking the exact opposite approach to that suggested by Sackett and Roth (1996). According to this recommendation, there are instances when organizations would want to use their high-adverse-impact tests in the first hurdle and allow proportionally slightly more applicants through. Then, organizations would use lower-adverse-impact assessments in the later hurdle and allow proportionately slightly fewer applicants through.

In summary, a multiple-hurdle approach to selection allows organizations to sequence assessments in order to minimize adverse impact and cost. This approach takes a noncompensatory view to selection, wherein high scores on all assessments are required to pass the selection process. Moreover, the theory behind multiple-hurdle approaches to selection is nonlinear; after the cutoff score, all participants are considered equal.

USING MULTIPLE COMBINATION METHODS

We discussed each of these techniques separately, almost as if organizations never combine these methods when making decisions about applicants. In reality, these techniques can be combined to help identify the best overall candidate. For example, let us assume that an organization uses three different assessments (e.g., cognitive ability test, physical skill assessment, and interview) to evaluate its job applicants. Pass scores could be identified separately for each assessment tool, and these assessment tools are then used in a multiple-hurdle fashion. However, what happens when multiple people score above all three pass scores? The quantitative score on each assessment can be combined using the compensatory approach. While the remaining applicants all have the requisite level of each latent skill or ability, how do we determine who to hire when multiple applicants survive all of the hurdles? The compensatory approach can be used to combine scores across all three predictors. This way strengths in one area can compensate for moderate weaknesses in another. However, true deficiencies in one area can never be compensated for by strengths in another, because individuals whose scores fall below a pass score are eliminated from the potential future employee pool by the multiple-hurdle technique. Thus, in applied settings, it is likely that the various techniques are used simultaneously to provide the optimal and most appropriate decisions regarding the applicants.

CONCLUSION

In this chapter, we discussed various techniques that have been developed to enable organizations to combine information from multiple assessments to form a holistic impression of their job applicants. Specifically, we discussed qualitative methods (aka clinical assessment) and several quantitative methods (e.g., compensatory, multiple-hurdle, Pareto-optimal). The strengths and weaknesses of these techniques, along with recommendations from the literature regarding best practices when using these methods, were noted.

REFERENCES

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382.

Bernardin, H. J., & Buckey, M. R. (1981). Strategies in rating training. *Academy of Management Review, 6*, 205–212.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores. *Organizational Research Methods, 10*, 689–709.

De Corte, W., Lievens, F., & Sackett, P. R. (2006). Predicting adverse impact and mean criterion performance in multistage selection. *Journal of Applied Psychology, 91*, 523–537.

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology, 92*, 1380–1393.

De Corte, W., Lievens, F., & Sackett, P. R. (2008). Validity and adverse impact potential of predictor composite formation. *International Journal of Selection and Assessment, 16*, 183–194.

De Corte, W., Sackett, P., & Lievens, F. (2010). Selecting predictor subsets: Considering validity and adverse impact. *International Journal of Selection and Assessment, 18*, 260–270.

De Soete, B., Lievens, F., & Druart, C. (2012). An update on the diversity—validity dilemma in personnel selection: A review. *Psychological Topics, 21*, 399–424.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical predictors: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Hattrup, K., & Rock, J. (2002). A comparison of predictor-based and criterion-based methods for weighting predictors to reduce adverse impact. *Applied H.R.M. Research, 7*, 22–38.

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology, 82*, 656–664.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1*, 333–342.

Hollenbeck, G. P. (2009). Executive selectin—What's right . . . and What's wrong. *Industrial and Organizational Psychology, 2*, 130–143.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial work and organizational psychology* (Vol. 1, pp. 233–377). London: Sage.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9*, 152–194.

Johnson, J. W., & Carter, G. W. (2010). Validating synthetic validation: Comparing traditional and synthetic validity coefficients. *Personnel Psychology, 6*, 755–795.

Kanfer, R., Wolf, M. B., Kantrowitz, T. M., & Ackerman, P. L. (2010). Ability and trait complex predictors of academic and job performance: A person-situation approach. *Applied Psychology, 59*, 40–69.

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*, 1060–1072.
Morris, S. B., Daisley, R. L., Wheeler, M., & Boyer, P. (2014). A meta-analysis of the relationship between individual assessments and job performance. *Journal of Applied Psychology, 100*, 5–20.
Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior and Human Decision Processes, 38*, 76–91.
Pyburn, K. M., Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61*, 143–151.
Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707–722.
Sackett, P. R., & Roth, L. (1996). Multi-state selection strategies: A Monte Carlo investigation of effects of performance and minority hiring. *Personnel Psychology, 49*, 549–572.
Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist, 56*, 302–318.
Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*, 735–746.
Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719–730.
Thornton, G. C., III., Hollenbeck, G. P., & Johnson, S. K. (2010). Selecting leaders: Executives and high potentials. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 823–840). New York, NY: Routledge/Taylor & Francis Group.
Weiner, I. B. (2003). The assessment process. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Vol. 10: Assessment psychology* (pp. 3–25). Hoboken, NJ: Wiley.
Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.

Downloaded By: 10.2.97.136 At: 16:24 26 Sep 2023; For: 9781315690193, chapter17, 10.4324/9781315690193-17