

PRACTICE FORUM

Addressing the so-called validity–diversity trade-off: Exploring the practicalities and legal defensibility of Pareto-optimization for reducing adverse impact within personnel selection

Deborah E. Rupp^{1,*}, Q. Chelsea Song², and Nicole Strah²

¹George Mason University and ²Purdue University

*Corresponding author. Email: drupp2@gmu.edu

(Received 01 August 2019; revised 30 October 2019; accepted 19 November 2019)

Abstract

It is necessary for personnel selection systems to be effective, fair, and legally appropriate. Sometimes these goals are complementary, whereas other times they conflict (leading to the so-called “validity-diversity dilemma”). In this practice forum, we trace the history and legality of proposed approaches for simultaneously maximizing job performance and diversity through personnel selection, leading to a review of a more recent method, the Pareto-optimization approach. We first describe the method at various levels of complexity and provide guidance (with examples) for implementing the technique in practice. Then, we review the potential points at which the method might be challenged legally and present defenses against those challenges. Finally, we conclude with practical tips for implementing Pareto-optimization within personnel selection.

Keywords personnel selection; Pareto-optimal weighting; adverse impact; validity-diversity; discrimination; race discrimination

“Effectiveness” within personnel selection is not a unitary construct. It refers to the extent to which an organization can successfully map job tasks onto necessary worker knowledge, skills, and abilities; measure these attributes; and recruit a deep applicant pool possessing these qualities (Gatewood et al., 2015). It also refers to an organization’s capability to mitigate legal risk, via compliance with equal employment opportunity laws (Equal Employment Opportunity Commission, 1978; Gutman et al., 2011). Finally, it refers to ensuring that those interacting with the selection process (both internal and external stakeholders) experience it as job-related and fair, and that the resulting hiring decisions promote the organization’s diversity and inclusion goals.

Unfortunately, organizations can face trade-offs when it comes to simultaneously attaining these inter-related yet at times competing goals (e.g., Pyburn et al., 2008). For example, whereas a firm may have access to a hiring tool that effectively predicts future job performance, the tool may not be compliant with legal guidance (e.g., the standards specified by the *Uniform Guidelines*; Equal Employment Opportunity Commission, 1978); it may be *seen* as disadvantaging minority groups (e.g., a physical ability test seen as unfair by women or disabled applicants); or it may cause legitimate adverse impact, where groups protected by legislation (e.g., Title VII of the Civil Rights Act of 1964) are systematically underselected. This is further complicated by the fact that the

All authors contributed equally to this article.

© Society for Industrial and Organizational Psychology, Inc. 2020.

nature of much equal employment opportunity (EEO) legislation in the United States prohibits “radical” forms of discrimination reduction, such as giving preferential treatment to minorities in efforts to increase workforce diversity (except under specific circumstances such as when court ordered or as part of a temporary voluntary settlement agreement).

As such, the field of industrial and organizational (I-O) psychology has expended considerable effort examining how these tradeoffs might be effectively balanced in order to maximize the aforementioned effectiveness criteria. To date, these efforts have received mixed success. Strategically integrating recruiting and selection efforts has shown some potential. For example, while general minority recruiting is unlikely to result in greater diversity of selected applicants if adverse impact-prone selection tools are used (Tam et al., 2004), there is preliminary support for qualification-based minority recruiting (Newman et al., 2014; Newman & Lyon, 2009). While more sophisticated psychometric innovations, such as banding (i.e., using standard errors to create score bands where applicants within a band are considered equal and then making diversity-related hires within a band), have shown practical potential, they have been viewed by the courts as legally non-compliant (e.g., Henle, 2004; Campion et al., 2001).

In this practice forum, we briefly trace these efforts, exploring both the technical aspects of the proposed methods, as well as the technicalities on which they have been called out within the court system. We then review a more recent method that has been proposed in the psychometric literature, Pareto-optimization, which seeks to simultaneously maximize the prediction of job performance and minimize adverse impact against legally protected groups. As we review below, whereas this method has promise and overcomes some of the pitfalls of previous methods, it has yet to be widely applied within organizations and, consequently, vetted in the court system through legal challenge. We seek to contribute to both the research on and the practice of equal opportunity employment by providing a user-friendly, yet thorough, description of the Pareto-optimization method. We also provide materials to aid in the application of the method (ranging from a practical description of how to use a publicly available Pareto-optimization tool, to an in-depth, technical description of how the method performs the estimations). Finally, we discuss how the method could potentially be challenged legally as violating Title VII and what an organization could present (both proactively and reactively) as a defense against such challenges. By demystifying Pareto-optimization, our hope is that this presentation, critique, and analysis will serve to encourage organizations to consider employing the method within their selection systems. Further, we hope to encourage discourse among I-O psychologists and legal professionals regarding its defensibility if challenged on legal grounds.

Strategies for simultaneously maximizing diversity and validity

As mentioned above, some selection procedures that strongly predict job performance¹ show systematic differences in scores across demographic groups such as race or gender (Bobko & Roth, 2013; Ployhart & Holtz, 2008), which hinder the goals of diversity and legal defensibility by selecting minority applicants at lower rates than majority applicants.² This problem has been labeled the *diversity-validity dilemma* (Pyburn et al., 2008). The most salient example of this dilemma stems from the use of cognitive ability tests for personnel selection, which have been shown to be among

¹Our general use of the term “performance” at this stage in the article refers to traditional measures of task performance. That being said, we acknowledge that performance is a complex and multidimensional construct (Cascio & Aguinis, 2011), and that the specific performance criteria employed can impact both validity estimates and selection decisions (and hence adverse impact). We return to this issue later in the article.

²Importantly, these differences in expected performance for different demographic groups often do not align with actual subgroup differences in assessed performance (McKay & McDaniel, 2006).

the strongest predictors of job performance (corrected $r = .51$; Schmidt & Hunter, 1998; Roth et al., 2011), but also demonstrate higher race-related subgroup differences compared to other common selection procedures (Roth et al., 2001).³

A number of strategies have been proposed for mitigating this dilemma. For example, a surge of work discussing the use of score banding began in the early 1990s (cf. Campion et al., 2001). This method involves the use of standard errors to create score bands. Within a score band, candidates are considered to have statistically equivalent scores and thus cannot be rank-ordered by score. As such, preferential selection based on demographic status is one strategy that can be used to select candidates within a band to increase the representation of women and minorities in the workforce, while minimizing any potential compromise to validity. Unfortunately, this strategy has faced two major hurdles in its implementation. The first is an issue of recruitment. If an organization is not successful in recruiting a diverse pool of highly qualified applicants to begin with, then top score bands could potentially lack diversity, thus constraining what banding can do to improve diversity (Newman et al., 2014). The second is an issue of legality. The technique has in fact been legally challenged on the grounds of discriminating against White applicants, and it has been argued that using demographic status as a selection criterion within a band could be viewed as violating EEO law (i.e., interpreted as a race/sex/etc.-based decision; Henle, 2004).⁴ Whereas other, more legally appropriate selection criteria could be used to choose candidates within a band, these alternative methods may do little to resolve validity-diversity tradeoffs (Cascio et al., 2010).

A second approach involves strategically choosing, ordering, and combining various predictors in a way that minimizes validity-diversity tradeoffs (Ployhart & Holtz, 2008). This might involve the use of a variety of (valid, non-overlapping) predictors instead of, or in combination with, general cognitive ability tests. These predictors include tests of more fine-grained cognitive facets (e.g., verbal ability, mathematical ability), noncognitive predictors (e.g., personality), and predictor measures that are not tests (e.g., interviews, assessment centers; Bobko et al., 1999; Sackett & Ellingson, 1997). Illustrating this, Finch et al. (2009) showed through a simulation study that using different combinations of predictors in different stages of a selection procedure can, to varying extents, impact both the amount of performance predicted by the selection system as well as the adverse impact of the selection system for minority applicants.⁵ Further, when job performance is widely conceptualized to include task performance and contextual performance (rather than task performance only), it can be effectively predicted by a wider range of noncognitive predictors (that have smaller subgroup differences; Hatrup et al., 1997).

Despite the promise of these various approaches, a systematic method for carrying out such “optimization” is lacking. Instead, the majority of previous efforts rely on labor-intensive, trial-and-error-based analyses, and have been more academic exercises (to explore what is possible) than clear, actionable solutions for practitioners to implement. A defined process that allows for a more systematic identification of how predictors should be weighted across specific contexts could reduce the guesswork required within selection decisions and allow greater precision in reaching organizational goals for employee performance and diversity. The *Pareto-optimal weighting technique* (or “Pareto-optimization”) has been offered as just such a process.⁶

³We note that a variety of explanatory mechanisms have been proposed to account for subgroup differences in cognitive ability tests. Though a thorough explication of these mechanisms is beyond the scope of this article, we recommend Outtz and Newman (2010) for a review. See also Cottrell et al. (2015) for recent empirical work on this issue.

⁴Although Henle (2004) notes that multiple district and circuit courts have ruled that diversity status can be used as a “plus” factor in addition to other factors used to select applicants from a band (e.g., education, job experience, training).

⁵Though, as we discuss in the Practical Recommendations section, individual predictors are still vulnerable to legal challenge.

⁶However, it is important to recognize that even this method requires some judgment calls on the part of the user. We return to this issue later in the article.

Pareto-optimization as a potential way forward

The application of Pareto-optimization in personnel selection was first introduced by De Corte et al. (2007). This technique, which borrows from the multiple-objective optimization literature in engineering and economics, was proposed specifically to allow for personnel selection decisions that simultaneously optimize both the expected job performance and the diversity of new hires. Previous work (e.g., De Corte et al., 2008, 2020; Song, Wee et al., 2017; Wee et al., 2014) has shown that the technique has the potential to improve diversity (i.e., increase the number of job offers extended to minority applicants) with no loss in expected job performance.

Pareto-optimization involves intervening on the *weights* applied to the items or predictors comprising the selection system. It can be compared to more common methods such as unit weighting, where all predictors are given equal weights, and regression weighting, where regression analysis is used to determine predictor weights that maximize the prediction of job performance. Pareto-optimization differs from these techniques in that the weights for each predictor are statistically determined to simultaneously optimize two (or more) criteria (e.g., job performance *and* diversity), as compared to optimizing only one criterion (e.g., job performance, such as when regression weights are estimated).

As an illustration, consider an organization that has identified three predictors for use in its selection system: a cognitive ability test, a conscientiousness (personality) measure, and a physical ability assessment. Screening candidates using these three tools results in a score on each predictor. *Unit weighting* involves taking the sum or average of the three predictor scores and rank-ordering candidates on the basis of the total or average scores. Unit weighting makes intuitive sense, and research has supported its effective use in making valid selection decisions (e.g., Einhorn & Hogarth, 1975), which has contributed to its popularity in practice (Bobko et al., 2007).

Regression weighting involves obtaining predictor scores from current employees, along with an accurate measure of their job performance (the criterion).⁷ The data are then used to fit a regression model to determine the weighting scheme that most likely maximizes (optimizes) criterion-related selection decisions. These regression weights are then used to calculate a weighted predictor composite score for each applicant, which allows for the rank-ordering of the applicants in a top-down selection process. Although regression weighting is more labor intensive and requires reliable performance data, its promise for maximizing criterion-related validity has also made it a popular technique within organizations (Bobko, 2001).

Pareto-optimal weighting is similar to regression weighting in that it also seeks “optimized” composite scores. However, it differs from regression weighting in that it aims to optimize two (or more) outcomes simultaneously (De Corte et al., 2007). For example, if an organization wants to simultaneously minimize certain subgroup differences/adverse impact and maximize job performance in new hires, Pareto-optimization could be utilized to derive a set of weights that provides the best possible solution (e.g., maximizing performance prediction at a predetermined threshold for adverse impact). That is, given a desired level of diversity among new hires, a maximum level of expected job performance can be obtained. Likewise, given a desired level of expected job performance, a maximum level of diversity among new hires can be obtained.

Continuing our example from above, where an organization has in place three selection predictors for a particular job (cognitive ability, conscientiousness, and physical ability), let us further

⁷Although performance as a psychological construct is multidimensional (Motowidlo & Kell, 2012), it is often treated as a unidimensional composite (and measured as basic task performance) within criterion-related validity studies and within weighting methods such as these. We discuss performance dimensionality in more detail later in this article.

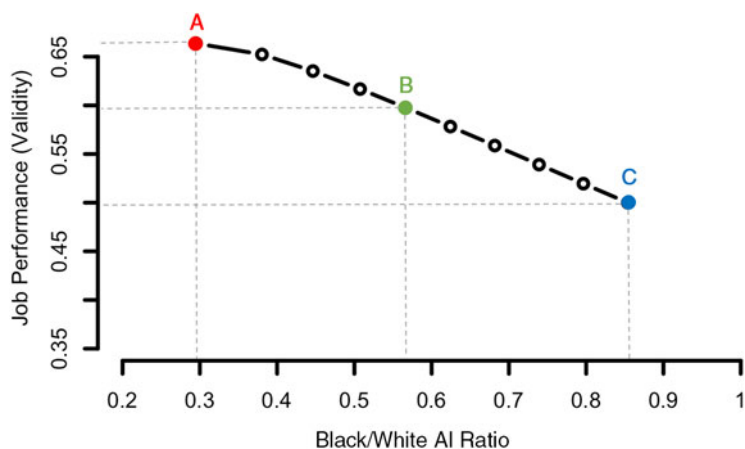


Figure 1. An example Pareto-optimal trade-off curve. Job performance (validity) is the correlation between the weighted predictor composite score and job performance score. Diversity of new hires is represented as the Black/White adverse impact (AI) ratio. Point A (red dot) represents the solution where job performance validity is maximal; Point C (blue dot) represents the solution where adverse impact ratio is maximal. Point B (green dot) represents the Pareto-optimal solution where the prediction of future job performance and the minimization of adverse impact against Black applicants are considered equally important.

assume that it has interest in reducing Black/White⁸ subgroup differences to the greatest extent possible without sacrificing its ability to effectively predict applicants' future job performance, and wishes to do so via Pareto-optimization. Figure 1 illustrates a potential Pareto-optimal tradeoff curve that could be used to attain this goal. The horizontal axis represents the adverse impact ratio (i.e., the selection rate of Black applicants relative to the selection rate of White applicants).⁹ Throughout our example, we reference the four-fifths or 80% rule (i.e., that the selection rate of one protected subgroup should not be less than 80% of the majority subgroup's selection rate; EEOC, 1978), while at the same time acknowledging it is not the only (or often best) way to demonstrate adverse impact.¹⁰ The vertical axis represents the possible range of criterion-related validity estimates (i.e., the varying levels at which selection composite scores correlate with job performance). Each point on the curve represents a set of potential predictor weights. Continuing our example with cognitive ability, conscientiousness, and physical ability as predictors, each point on the Pareto-optimal curve represents a set of three predictor weights, one for each of the three predictors. The negative slope of the Pareto-optimal curve illustrates the diversity-performance tradeoff. If the curve is steep, it means the organization must sacrifice a

⁸Note that we have specified a specific subgroup contrast (i.e., Black/White differences), as required by case law and EEOC guidance (though there are some instances of agencies lumping groups together to create larger sample sizes). The necessity to make specific contrasts such as these highlights a complexity of the Pareto-optimization method: that "optimal" weights may differ depending on what groups are being contrasted (e.g., Black/White; men/women), as well as what performance criterion is applied. We return to this issue in the Practical Recommendations section.

⁹Adverse impact ratio is a commonly used diversity criterion in Pareto-optimization studies (e.g., De Corte, 2014; De Corte et al., 2007; 2011; Druart & De Corte, 2012a; Song, Wee, & Newman, 2017; Wee et al., 2014). However, other diversity criteria could be used when applying Pareto-optimization. For example, De Corte et al. (2010, 2020) used minority selection rate as the diversity criterion. Other potential criteria could include Z-test estimates, as well as (a) the number of minority applicants selected given a specific overall selection ratio and (b) predictor-composite subgroup difference (see Newman et al., 2007). Although possible, incorporating chi-square, Fisher's exact, or Lancaster's mid-P test estimates is more complex, in that for these tests, the criteria for statistical significance vary by sample.

¹⁰That is, other methods have been called for in some circumstances, and legal precedent suggests using a method most appropriate for a given sample (see Gutman et al., 2017; Morris, 2017; Oswald et al., 2017). Whereas we use the 80% or four-fifths rule throughout our example, as we mentioned in our previous footnote, other statistical estimates of adverse impact could be applied as well.

large amount of expected job performance to gain a small decrease in adverse impact. On the other hand, if the curve is flat, it means that the organization would not have to give up much expected job performance to get a large payoff in terms of adverse impact reduction.

Organizations can effectively use such curves to determine the specific weighting scheme where, to the extent it is simultaneously possible, certain subgroup differences are minimized and job performance validity is maximized. For example, the weighting scheme at Point A in Figure 1 would provide maximal job performance and high adverse impact (where the selection rate of Black applicants is far less than 80% that of White applicants). This would be akin to regression weighting, where optimization occurs around a single criterion only (job performance). Similarly, Point C represents the weighting scheme that minimizes the difference in Black/White selection rates with no consideration of job performance validity. In contrast, Point B represents a weighting scheme where the prediction of future job performance and the minimization of adverse impact against Black applicants are considered equally important.

Applying Pareto-optimal weighting in the real world

Although the psychometric/empirical research on Pareto-optimization may seem “statistics-heavy” as an actionable tool for use in organization, its application is quite straightforward. Below, we provide three general levels of detail to aid in the implementation of Pareto-optimization: (a) the steps human resource (HR) practitioners would need to take to collect/obtain data from current employees to estimate Pareto-optimal solutions; (b) the steps that would be taken once this information is collected to generate the Pareto-optimal curve and associated predictor weights; and (c) the technical details pertaining to how exactly the weights are estimated. We expect that different readers (HR practitioners, psychometric consultants, attorneys, expert witnesses) may find different aspects of these descriptions useful.

Table 1 provides a step-by-step guide for collecting the information needed to compute the input statistics that feed the Pareto-optimization algorithm. The first step involves the collection of predictor (e.g., scores from cognitive ability tests, personality measures, physical ability tests) and criterion (e.g., job performance and adverse impact) data from existing employees.¹¹ Using these data, in Step 2, one can compute (a) predictor intercorrelations, (b) job performance criterion validity estimates for each predictor (i.e., the correlation between each predictor score and job performance ratings), and (c) subgroup differences on each predictor (*d*; the standardized group-mean difference between two groups; e.g., Black/White differences, or differences between men and women). This information is then used in Step 3 as input into the Pareto-optimization algorithm to obtain the predictor weights and Pareto curve.

There are at least three tools that can be used to carry out Pareto-optimization in personnel selection: (a) a FORTRAN program, TROFSS (De Corte et al., 2007), (b) an R package, “ParetoR” (Song, Wee, & Newman, 2017), and (c) a click-and-point web application, ParetoR Shiny app (Song, Wee, & Newman, 2017). For each of these tools, users input the data and estimates described above in order to generate (1) the Pareto-optimal predictor weights, (2) criterion solutions (i.e., job

¹¹Because incumbent (rather than applicant) data are used to estimate Pareto-optimal weights, it is important to, when possible, cross-validate the estimated weights with applicant data. Using Monte-Carlo simulation, Song, Wee, and Newman (2017) found that, when the Pareto-optimal weights were applied to an applicant sample, both expected diversity and job performance outcomes decreased (i.e., diversity and validity shrinkage occurred), especially when the incumbent sample was small (e.g., smaller than 100). Nonetheless, the Pareto-optimal solution still outperformed common unit weighted solutions, even after taking into account shrinkage. Although Monte-Carlo simulation is informative, it cannot capture all aspects of a selection setting (e.g., applicant characteristics, practical considerations). Thus, pilot testing is highly suggested when possible (where, prior to use for selection, the weights are applied to an applicant sample [without contributing to hiring decisions] and the performance of those hired [using existing methods] is used to evaluate the effectiveness of the weights for predicting performance and minimizing adverse impact). Such pilot trials can also reveal issues related to performance reliability and other “pipeline” issues.

Table 1. A step-by-step guide for implementing Pareto-optimization within personnel selection

<p>Step 1. Collect data from current employees</p> <ul style="list-style-type: none"> a) Predictor scores (e.g., scores from cognitive ability tests, personality measures) b) Job performance scores (e.g., job performance ratings) c) Demographic information (e.g., gender, race)
<p>Step 2. Compute input statistics for Pareto-optimization</p> <p>Using data collected in Step 1, compute input statistics for the Pareto-optimization algorithm:</p> <ul style="list-style-type: none"> a) <i>Predictor intercorrelations</i> (correct for range restriction to reflect the predictor intercorrelations in the applicant sample) b) The <i>criterion-related validity</i> of each predictor (i.e., the correlation between each predictor score and job performance scores; correct for range restriction to reflect the criterion-related validity of the predictors in the applicant sample) c) <i>Subgroup differences for predictor scores</i> (<i>d</i>; the standardized group-mean difference between two groups, e.g., Black/White differences, or differences between men and women) <p>In addition, the following statistics that characterize the selection scenario need to be specified and/or calculated:</p> <ul style="list-style-type: none"> d) Selection ratio: Target proportion of applicants to be selected from all the applicants $\text{Selection ratio (SR)} = (\text{number of applicants selected}) / (\text{total number of applicants})$ e) Proportion of minority applicants: The estimated proportion of minority vs. total applicants (e.g., using minority population from labor statistics) among all the applicants comprising the specific subgroup of interest (e.g., Blacks) $\text{Prop. of minority app.} = (\text{number of minority applicants}) / (\text{total number of applicants})$
<p>Step 3. Obtain the Pareto-optimal predictor weights</p> <ul style="list-style-type: none"> a) Using the input statistics from Step 2, estimate Pareto-optimal weights using one of (at least) three tools (see the Appendix for technical details of Pareto-optimization): <ul style="list-style-type: none"> ○ TROFSS FORTRAN program (see De Corte <i>et al.</i>, 2007) https://users.ugent.be/~wdecorte/software.html ○ ParetoR R package (see Song, Wee, & Newman, 2017) https://github.com/Diversity-ParetoOptimal/ParetoR ○ ParetoR R Shiny web application (see Song, Wee, & Newman, 2017) https://qchelseasong.shinyapps.io/ParetoR/ b) Choose a set of Pareto-optimal weights from the resulting Pareto curve that provides desired diversity and job performance outcomes c) When possible, pilot test weights on applicants selected via current system
<p>Step 4. Implement the Pareto-optimal predictor weights in selection of job applicants</p> <ul style="list-style-type: none"> a) Collect predictor scores from the job applicants b) For each applicant, calculate weighted composite scores using the Pareto-optimal weights from Step 3 c) Make hiring decisions based on (or partially based on) the Pareto composite scores

performance validity and AI ratios), and (3) the Pareto-optimal trade-off curve. The Appendix provides the technical details pertaining to how the Pareto-optimal weights are generated.

Figure 2 shows a screenshot of the ParetoR Shiny app. The web application consists of three parts: (1) “Input” (red box, left, and Figure 3); (2) “Output: Plots” (green box, top right, and Figure 4; and (3) “Output: Table” (blue box, bottom right, and Figure 5). To start, users specify as input (a) the selection ratio (the expected percentage of applicants who will be extended job offers), (b) the expected proportion of “minority”¹² applicants (following our example, this would be the expected proportion of Black applicants),¹³ (c) the number of predictors, (d) the predictor-criterion correlation matrix (obtained from the incumbent sample), and (e) predictor subgroup differences (obtained from the incumbent sample; following our example, this would be Black/White subgroup differences). Once this information is entered, the “Get Pareto-Optimal Solution!” button in the “Input” panel is selected.

¹²We use this term loosely, recognizing that a protected subgroup need not be in the minority to receive Title VII protections.

¹³This is commonly operationalized as the proportion of minority applicants in the population (e.g., De Corte *et al.*, 2020). For example, Potosky *et al.* (2005) set this proportion to correspond to the workforce composition as reported in 2000 by the Bureau of Labor Statistics (e.g., 88.1% White, 11.9% African American).

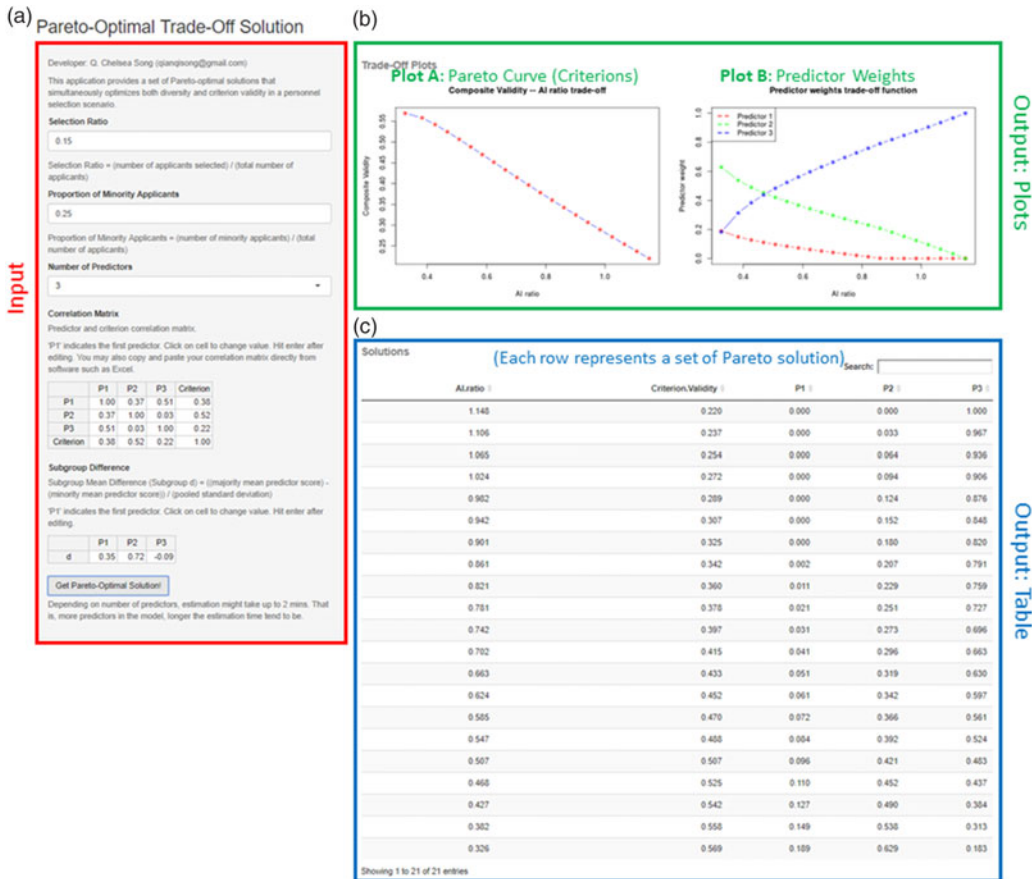


Figure 2. An example application of Pareto-optimization using the ParetoR Shiny app web application (<https://qchelseasong.shinyapps.io/ParetoR/>). The “Input” box (red box, left) shows the input panel in which users will specify the following input data: (a) selection ratio, (b) expected proportion of minority applicants, (c) number of predictors, (d) predictor-criterion correlation matrix, and (e) subgroup differences (see Table 1, Step 2 for details). The “Output: Plots” box (green box, top right) shows two output plots: (1) Plot A: Pareto-optimal trade-off curve [i.e., job performance validity (vertical axis) and AI ratio (horizontal axis) trade-off curve], similar to Figure 1, each point (Pareto point) on the trade-off curve represents a set of predictor weights; (2) Plot B: predictor weights across different Pareto points (the three predictors at each point correspond to the Pareto point in Plot A). The “Output: Table” box (blue box, bottom right) shows the AI ratio, job performance criterion validity, and predictor weights corresponding to each Pareto point (each row). Expanded views of each of the parts of this figure are shown in Figures 3–5.

The Pareto-optimal solutions will be displayed in the “Output Plots” and “Output: Table” sections on the right. Plot A in the “Output: Plots” section (Figure 2, green box, and Figure 4) is a Pareto curve, similar to Figure 1. The vertical axis in Plot A displays criterion-related validities (performance outcomes), whereas the horizontal axis displays adverse impact ratios (diversity outcomes). The Shiny app provides 21 Pareto points, or 21 evenly spaced solutions. Each point (Pareto points) on the trade-off curve represents a set of predictor weights (three predictors in our ongoing example) that simultaneously optimize both job performance (criterion validity) and the Black/White adverse impact ratio. In the example, as the user examines the curve from left to right, the sets of predictors increasingly provide less job performance criterion validity and more favorable Black/White adverse impact ratios.

Plot B presents the predictor weights across different Pareto points. In the example, more weight is given to Predictor 2 when job performance criterion validity is maximal (and the Black/White adverse impact ratio is not maximal), and Predictor 3 is weighted more when the

Input

Developer: Q. Chelsea Song (qianqisong@gmail.com)

This application provides a set of Pareto-optimal solutions that simultaneously optimizes both diversity and criterion validity in a personnel selection scenario.

Selection Ratio

Selection Ratio = (number of applicants selected) / (total number of applicants)

Proportion of Minority Applicants

Proportion of Minority Applicants = (number of minority applicants) / (total number of applicants)

Number of Predictors

Correlation Matrix

Predictor and criterion correlation matrix.

'P1' indicates the first predictor. Click on cell to change value. Hit enter after editing. You may also copy and paste your correlation matrix directly from software such as Excel.

	P1	P2	P3	Criterion
P1	1.00	0.37	0.51	0.38
P2	0.37	1.00	0.03	0.52
P3	0.51	0.03	1.00	0.22
Criterion	0.38	0.52	0.22	1.00

Subgroup Difference

Subgroup Mean Difference (Subgroup d) = ((majority mean predictor score) - (minority mean predictor score)) / (pooled standard deviation)

'P1' indicates the first predictor. Click on cell to change value. Hit enter after editing.

	P1	P2	P3
d	0.35	0.72	-0.09

Depending on number of predictors, estimation might take up to 2 mins. That is, more predictors in the model, longer the estimation time tend to be.

Figure 3. An expanded view of Figure 2: "Input."

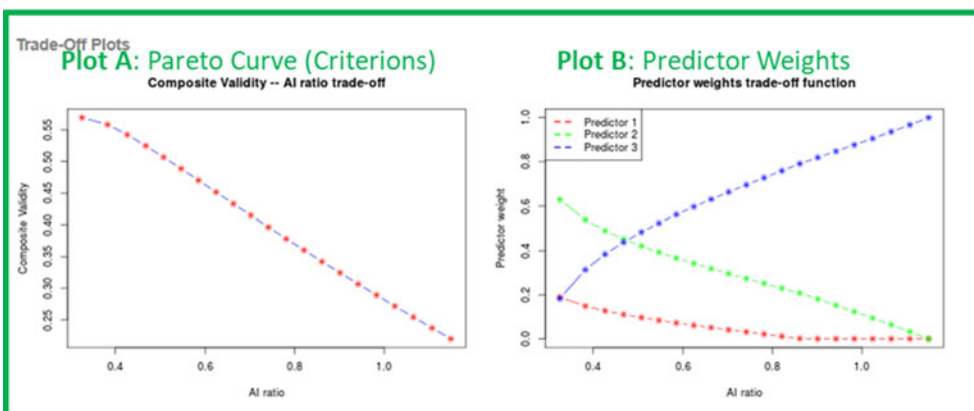


Figure 4. An expanded view of Figure 2: "Output: Plots."

Solutions

(Each row represents a set of Pareto solution)

Search:

AI.ratio	Criterion.Validity	P1	P2	P3
1.148	0.220	0.000	0.000	1.000
1.106	0.237	0.000	0.033	0.967
1.065	0.254	0.000	0.064	0.936
1.024	0.272	0.000	0.094	0.906
0.982	0.289	0.000	0.124	0.876
0.942	0.307	0.000	0.152	0.846
0.901	0.325	0.000	0.180	0.820
0.861	0.342	0.002	0.207	0.791
→ 0.821	0.360	0.011	0.229	0.759
0.781	0.378	0.021	0.251	0.727
0.742	0.397	0.031	0.273	0.696
0.702	0.415	0.041	0.296	0.663
0.663	0.433	0.051	0.319	0.630
0.624	0.452	0.061	0.342	0.597
0.585	0.470	0.072	0.366	0.561
0.547	0.488	0.084	0.392	0.524
0.507	0.507	0.096	0.421	0.483
0.468	0.525	0.110	0.452	0.437
0.427	0.542	0.127	0.490	0.384
0.382	0.558	0.149	0.538	0.313
0.326	0.569	0.189	0.629	0.183

Showing 1 to 21 of 21 entries

Output: Table

Figure 5. An expanded view of Figure 2: “Output: Table.”

Black/White adverse impact ratio is maximal (and job performance is not maximal). This is because Predictor 2 is the strongest predictor of job performance ($r = .52$; see “Correlation Matrix” in the “Input” panel) but is also affiliated with the highest Black/White subgroup d ($d = .72$; see “Subgroup Difference” in the “Input” panel). In contrast, Predictor 3 is the weakest predictor of job performance ($r = .22$, see “Correlation Matrix” in the “Input” panel) but is also affiliated with the lowest Black/White subgroup d ($d = -.09$, see “Subgroup Difference” in the “Input” panel).

The “Output: Table” box (Figure 2, blue box, bottom right; Figure 5 shows an expanded view) presents the specific adverse impact ratio, job performance criterion validity, and predictor weights corresponding to each Pareto point (each row), plotted in the “Output: Plots” section. Based on the selection outcomes (i.e., adverse impact ratio, job performance criterion validity) listed in the table, users can select a combination of (in this case three) predictor weights that lead to their preferred outcome (out of the 21 sets of predictor weights). For example, an organization might choose the solution that results in a Black/White AI ratio of .82 and job performance criterion validity of .36 (Figure 5, the row with an arrow). This is the solution out of the 21 Pareto-optimal solutions that provides the highest job performance criterion validity for an adverse impact ratio that is greater than .80 (the four-fifths rule often referred to in court and within the *Uniform Guidelines*¹⁴).

¹⁴Though we note again that many courts and agencies have since expressed disfavor for the four-fifths rule, instead favoring methods of statistical significance (Gutman et al., 2017; Oswald et al., 2017).

In this example, if these were the primary subgroups of interest, and if compliance to the fourth-fifths rule was a goal (along with the best possible criterion-related validity given this goal), in subsequent selection processes, users would give the weights of .01, .23, and .76 for Predictors 1, 2, and 3, respectively.¹⁵ However, it may be the case that they want to also consider the Pareto curve for other subgroup differences (e.g., women/men). Our example provides a situation where this might very well be the case, and highlights one complexity of Pareto-optimization: simultaneously considering multiple subgroups. Whereas the use of cognitive ability tests has historically shown adverse impact against racial minority groups, the use of physical ability tests has historically shown adverse impact against women as compared to men. Thus, generating the Pareto curve considering sex-based subgroups will likely produce a different set of “optimal” weights. This highlights the need for users’ specific selection goals to guide the analyses they choose to carry out. It also suggests considering ahead of time how the opposing “optimal” weighting schemes for different subgroup comparisons will be reconciled when making final decisions about the scoring and weighting of predictors. We discuss this issue further in the Practical Recommendations section.

Considerations of the legal defensibility of Pareto-optimization

Whereas, statistically speaking, Pareto-optimization offers a promising solution for dealing with diversity-validity tradeoffs, this does not necessarily mean that it has the power to hold up to legal scrutiny. This was the very issue our field faced following promising research on the banding technique for personnel selection (Henle, 2004). That is, although the method offered a solution for increasing minority selection rates using predictors with strong job performance criterion validity, the courts determined that giving preference to minority group members within a band is illegal, even though their selection scores can be considered equivalent to majority group members in the same band. As such, it is important that we identify ways Pareto-optimization might be legally challenged, alongside possible defenses organizations might offer in response to such challenges (as well as proactive actions organizations might take to avoid such scrutiny).

As was the case with banding, given that the use of Pareto-optimization is aimed at increasing the representation of underrepresented groups in the workplace (or at least decreasing the reduction of minority hiring rates that might be caused by using validity-based regression weights alone), challenges claiming traditional adverse impact (against minority members) may be unlikely. What could occur, however, are challenges in the form of “majority” group members (e.g., Whites, men¹⁶) claiming so-called “reverse discrimination” in the form of (intentional) disparate treatment (since considering diversity is explicitly and purposely a goal of the method). Such claims are not unprecedented in general, as evidenced by the cases discussed below. That being said, an extensive search of case law and legal discourse revealed no reports of Pareto-optimization having been legally challenged in any way. Our search also did not reveal any discussion of the method as explicitly legally defensible, providing what we consider to be a clean slate for considering the legality of Pareto-optimization.

¹⁵If decision makers choose a weighting scheme where certain predictors are not weighted substantially above 0 (e.g., in this case, Predictor 1 is only weighted at .01), decision makers might choose to omit these predictors for practicality. Though psychometrically, including versus not including Predictor 1 will make little difference in who is extended offers (and its exclusion could reduce costs), legal precedent (e.g., *Ricci v. DeStefano*, 2009) suggests that the safest choice is to only omit predictors *prior* to implementing the selection system. This should occur as a part of the validation/design phase. Care should be taken when adapting a selection system subsequent to the screening of applicants. See the legal defensibility section below for more detail.

¹⁶... or potentially other “minority” or underrepresented groups who were not considered in the computation of the Pareto-optimal weights, and/or whose adverse impact odds were thought to have increased as a result of the focal group’s odds decreasing. See the Practical Recommendations sections for more discussion of this issue.

Challenge: Demographic information used to determine weights

One specific legal challenge that might be made to the application of the Pareto-optimization is that, since demographic information is used to compute the weights, this method might violate Title VII of the Civil Rights Act (which prohibits selection decisions on the basis of race, color, religion, sex, or national origin). This would be a *disparate treatment* claim, where an argument is made that the organization *intentionally* took protected group status into consideration when making selection decisions. One defense against such a claim is that the method does not consider any demographic information from the *applicant* pool itself; nor does it consider demographic information of any *individual* job incumbent. Rather, the only demographic information used to determine the Pareto-optimal weights are *group-level* subgroup differences among *current* employees on predictor and criterion variables (De Corte et al., 2007).

Challenge: Method advantages underrepresented groups

A second challenge that might be brought forth concerning the use of Pareto-optimization is whether the method violates Title VII by virtue of creating an advantage for underrepresented groups (e.g., women, racial minorities). The argument would be that, by incorporating information on incumbent subgroup differences alongside criterion-related validity estimates to create the Pareto curve, organizations are consciously choosing to sacrifice some level of validity (“performance”) in order to gain diversity, which could be interpreted as (“illegal”) preferential treatment (disparate treatment) in favor of “minority” group members.

In *Hayden v. County of Nassau* (1999), Nassau was under a consent decree with the Department of Justice. Motivated by a goal to create as little adverse impact in selection as possible, Nassau consciously chose to use only nine sections of a 25-section exam that was taken by over 25,000 applicants (in order to retain as much criterion-related validity as possible while preventing as much adverse impact as possible). Importantly, when choosing the test sections on which the company was going to actually consider applicants’ scores for the selection process, Nassau rejected a different subset of exam components that led to a better minority selection ratio but worse criterion validity for performance. The Second Circuit Court ruled that, although the test redesign took race into account, it was scored in a race-neutral fashion and therefore was acceptable. Further, the court noted that attempts to alleviate adverse impact for a minority group are not synonymous with intentionally discriminating against the majority group.

This ruling, however, is interpreted by some as having been overturned by the Supreme Court ruling in *Ricci v. DeStefano* (2009)—except for instances when a consent decree is in place (Gutman et al., 2011). In this case, the New Haven Fire Department was using an exam as part of a selection process for promotions to the ranks of lieutenant and captain. After administering the test, New Haven feared it would face an adverse impact liability given that the results would have led to no Black candidates being promoted. Consequently, they invalidated the test results and used an alternative approach to make promotion decisions. Majority group candidates then filed a disparate treatment claim, arguing that race was used as a factor in determining what predictor to use (and not use) in making promotion decisions. In a split vote, the Supreme Court supported this reasoning in its ruling, setting the standard that “before an employer can engage in intentional discrimination for the asserted purpose of avoiding or remedying an unintentional, disparate impact, the employer must have a strong basis in evidence to believe it will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory action” (*Ricci v. DeStefano*, 2009). In a dissent supported by three other justices, Justice Ginsburg noted the heavy burden on an organization to show a strong basis in evidence when taking actions aimed at mitigating adverse impact, stating, “This court has repeatedly emphasized that the statute ‘should not be read to thwart’ efforts at voluntary compliance The strong-basis-in-evidence standard,

however as barely described in general, and cavalierly applied in this case, makes voluntary compliance a hazardous venture” (*Ricci v. DeStefano*, 2009, Ginsburg dissenting opinion).

This ruling might, at first blush, be interpreted as precedent against using Pareto-optimized weights, as New Haven did explicitly take adverse impact—and therefore race—into account when deciding whether and how to use test scores. Indeed, case law recommends great care be taken in how predictors are chosen and combined, and legal analyses suggest that *Ricci* “highlights the tension between the requirement to seek alternatives with less (or no) adverse impact and disparate treatment rules that permit non-minorities to claim disparate treatment if alternatives are used. Therefore, employers . . . may be sued for either using or not using alternatives” (Gutman *et al.*, 2011, p. 151).

That having been said, there are a number of differences in what occurred at Nassau County and New Haven compared to the scenario we describe in the current article, where an organization proactively decides to employ Pareto-optimal weights to address validity-diversity tradeoffs. First, in our scenario, tests have not already been administered to job applicants in the way they were at Nassau and New Haven. That is, *incumbent* data are used in determining how predictors will be weighted *a priori*. In this way, consistent with best practice, test scores are piloted and validated in advance, and in determining if and how they will be used/weighted, considered for their ability to both predict variance in job performance and minimize adverse impact. This is consistent with the holistic view of validity that professional standards have come to embrace (e.g., the Society for Industrial and Organizational Psychology, 2018), which emphasizes both prediction *and* fairness as evidence for the effectiveness of selection tests.

Second, much of the field of personnel selection is predicated on making smart decisions in determining what predictors to use and how predictor scores will be combined (e.g., Sackett & Lievens, 2008). It is considered “industry standard” to identify predictors that can be supported by job analysis, and, in deciding how to use these predictors to consider other factors, such as the potential for adverse impact, cost, and so forth (Ployhart & Holtz, 2008). The use of personality assessment within selection contexts is often the result of such an approach. The only difference with the use of Pareto-optimal weights is that these decisions are more systematically quantified *in order to more effectively make such decisions*. Again, considering the potential for subgroup differences within the context of a validation study is completely consistent with professional standards (e.g., the *Principles for the Validation and Use of Personnel Selection Procedures* [SIOP, 2018]; the *Uniform Guidelines on Employee Selection Procedures* [EEOC, 1978]).

Third, the use of Pareto-optimal weights does not necessarily disadvantage majority applicants. That is, no within-group adjustments are made; rather, an additional criterion (adverse impact reduction) is applied as implementation/combination strategies are considered. Further, as explained above, in using the Pareto curve to choose weights, the organization can decide what thresholds to use in terms of both job performance validity and adverse impact. Many psychometric discussions of this method (e.g., Sackett *et al.*, 2010) use examples that set the adverse impact ratio to .80 in order to reflect the four-fifths standard often used by the courts. However, the adverse impact ratio does not have to be set at exactly .80 (and as mentioned in the footnotes above, alternative metrics to the adverse impact ratio could be used here as well). Rather, an organization can look to the Pareto curve with the goal of improving diversity (i.e., decreasing the adverse impact ratio) in ways that do not sacrifice job performance validity. The extent to which an organization will be able to attain this goal is, of course, dependent on the context itself (e.g., the predictors used, the presence of subgroup differences, the selection ratio, the expected proportion of minority applicants in the applicant pool).

Challenge: Does the method hold up to the Daubert standards?

A third way in which the legality of the use of Pareto-optimization might be questioned is related to the Daubert standard for the admissibility of expert testimony (originally established in *Daubert v. Merrell Dow Pharmaceuticals Inc.*, 1993). This standard has often been used to evaluate a

scientific or technical method that is argued to be rigorous and acceptable by a defendant or plaintiff's expert witness. There are five illustrative factors (criteria) used within legal contexts to support a method: (1) it can be or has been tested previously, (2) it has been published in peer-reviewed outlets, (3) its known or potential level of imprecision is acceptable, (4) there is evidence that a scientific community accepts the method (i.e., the method has widespread acceptance within the community and is not viewed with a large amount of skepticism), and (5) the method will be judged by the courts based on its inherent characteristics as opposed to the conclusions of the analysis.

The Pareto-optimal weighting technique fares well when held up against each of these criteria. First, this method has been tested in several peer-reviewed articles (summarized in Table 2), meeting criteria (1) and (2). The top section of Table 2 reviews the peer-reviewed articles and presentations that have both introduced and tested the validity of the technique. As an example, De Corte et al. (2008) and Wee et al. (2014) demonstrated how the Pareto-optimal method could simultaneously provide improved job performance and diversity outcomes as compared to both unit weighting and regression weighting (methods that have been widely accepted within the court system; see *Black Law Enforcement Officers Assoc. v. Akron*, 1986; Kaye, 2001; *Reynolds v. Ala. DOT*, 2003).

With regard to criterion (3) involving accuracy, when carrying out Pareto-optimization, confidence intervals can be created around an expected outcome (e.g., expected job performance criterion validity given a particular AI ratio) in order to take into consideration imprecision with regard to the data and measures.¹⁷ Further, Song, Wee, and Newman (2017) examined the cross-sample validity of Pareto-optimal weights. Cross-sample validity refers to the extent to which an optimal weighting solution for one calibration sample (e.g., the incumbent sample on which a selection system's weighting scheme was established) similarly predicts the outcomes of interest for a second validation sample (e.g., the applicant sample on which a selection weighting scheme could be used). As optimization methods such as Pareto-optimization aim to maximize their model fit in the calibration sample, the resulting weights tend to overfit, leading to smaller predictive validity in the validation sample as compared to the calibration sample (i.e., validity shrinkage). Song and colleagues demonstrated that when the calibration sample is sufficiently large (e.g., in their study, 100 participants for a set of cognitive selection predictors examined in their article), Pareto-optimization outperforms unit weighting methods in terms of maximizing both performance validity and diversity (represented by the adverse impact ratio), even after accounting for the possible loss in predictive validity in the validation sample.

With regard to criterion (4), Pareto-optimal weighting seems to have gleaned an acceptable level of support by the scientific community. Table 2 demonstrates that this method has been extensively discussed in academic writing within the applied psychological testing community. Not only has the method been explicitly examined in several empirical peer-reviewed articles (e.g., De Corte et al., 2007, 2008; Druart & De Corte, 2012a, 2012b; Wee et al., 2014) and commentaries of these articles (e.g., Kehoe, 2008; Potosky et al., 2008; Sackett & Lievens, 2008), but many reviews and book chapters (e.g., Aiken & Hanges, 2017; Oswald et al., 2014; Russell et al., 2014; Sackett et al., 2010) have discussed the method as a promising development within personnel selection contexts.

Finally, criterion (5) requires that the method can be judged by the courts based on its inherent characteristics (as opposed to the consequences of implementing the method in a particular context). Although we are unaware of any examples of the method being challenged and therefore judged by the courts, we believe that the summaries and arguments provided in this article along with the expansive work listed in Table 2 demonstrate the inherent viability of this method within selection systems (and thus the opportunity to judge its characteristics without bias).

¹⁷See Lopes et al. (2016), Naves et al. (2017), and Nimmegeers et al. (2016) for examples of how confidence intervals have been estimated for Pareto-optimal outcomes in other contexts.

Table 2. Summary of papers and presentations on the use of Pareto-optimization in personnel selection

Authors	Year	Brief summary	Evaluation
<i>Research on Pareto-optimization</i>			
De Corte, Lievens, & Sackett	2007	[Journal article] Introduced the use of Pareto-weighting method in personnel selection to enhance diversity in hiring and reduce adverse impact.	Positive
De Corte, Lievens, & Sackett	2008	[Journal article] Using a combination of cognitive and noncognitive predictors, demonstrated that the Pareto-optimal solution could provide better diversity and job performance outcomes as compared to regression weighting.	Positive
Kehoe	2008	[Commentary] Offered an organizational perspective on Pareto-optimality of validity-diversity trade-off. It concluded that Pareto-optimality allows organizations to address the question of decision making about selection quality and diversity from a broader context.	Positive
Potosky, Bobko, & Roth	2008	[Commentary] Compared the Pareto-optimal weighting approach to regression weighting approach to reduce adverse impact, and pointed out the importance of an organization's decision to improve diversity in hiring at possible cost of predicted job performance in resolving the diversity-performance dilemma.	Neutral
Sackett, De Corte, & Lievens	2008	[Commentary] Emphasized that the Pareto-optimal method is a complementary and not a competitive alternative for solving the trade-off between selection quality and adverse impact.	Positive
De Corte, Sackett, & Lievens	2010	[Journal article] Proposed a procedure for designing Pareto-optimal selection systems that systematically considers validity, adverse impact and constraints on the number of predictors that can be included in an operational selection system.	Positive
Druart & De Corte	2010	[SIOP poster] Expanded the use of the Pareto-optimal method to aid classification decisions, where one applicant pool was considered for multiple job positions simultaneously. Introduced a computer program that achieves optimal trade-offs between classification efficiency and diversity in a classification context.	Positive
Sackett, De Corte, & Lievens	2010	[Book chapter] Described Pareto-optimal weighting as a decision aid for adverse impact planning that permits examining trade-offs between adverse impact and mean criterion performance. Provided illustrative examples.	Positive
Tsang	2010	[Dissertation] Using the Pareto-optimal method, examined and demonstrated validity-diversity trade-off for two kinds of job performance: task performance and contextual performance. Results showed that reducing adverse impact required a greater validity trade-off for task performance than contextual performance.	Positive
De Corte, Sackett, & Lievens	2011	[Journal article] Presented an analytic method for designing Pareto-optimal selection systems that assists on six major selection design issues: (1) the predictor subset, (2) the selection rule, (3) the selection staging, (4) the predictor sequencing, (5) the predictor weighting, and (6) the stage retention decision issue.	Positive
Druart & De Corte	2012a	[Journal article] Described how the Pareto-optimal method can be applied to make optimal selection decisions when the applicants are allowed to simultaneously apply to one or more open positions. Concluded that the Pareto-optimal method can lead to substantive contributions in considering the quality-diversity dilemma in a complex selection context.	Positive

(Continued)

Table 2. (Continued)

Authors	Year	Brief summary	Evaluation
Druart & De Corte	2012b	[Journal article] Provided an example application of the Pareto-optimal method in complex selection situations where the applicants are allowed to simultaneously apply to multiple positions. Demonstrated the importance of applying the appropriate selection format (i.e., either the simple or the complex format) when exploring the Pareto-optimal solutions for planned selections.	Positive
De Soete, Lievens, & Druart	2012	[Review] Provided a review of the so-called diversity-validity dilemma and described the Pareto-optimal method as an effective method that uses a statistical approach to combine scores.	Positive
De Soete, Lievens, & Druart	2013	[Review] Provided an evidence-based overview of the effectiveness of six strategies for dealing with the so-called diversity-validity dilemma. Concluded that the Pareto-optimal method is one of the methods that holds the most promise to alleviate the dilemma.	Positive
De Corte	2014	[Journal article] Presented analytical methods to predict the Pareto-optimal outcomes of single- and multistage selection processes when applicant pools are small and heterogeneous. Presented computational tools for estimating the expected quality and diversity outcomes in single- and multistage selection settings.	Positive
Wee, Newman, & Joseph	2014	[Journal article] Examined the performance of Pareto-optimization when cognitive subtest scores were used as predictors in selection. Results showed that Pareto-optimization could lead to substantial improvement in diversity hiring, without decrements in expected future performance compared to unit weighted solution.	Positive
Lievens	2015	[Commentary] Discussed Pareto-optimal weighting as an example of mechanical combination to make selection decisions (which is preferred over judgment-based combinations).	Positive
Song, Wee, & Newman	2016	[SIOP symposium] Evaluated the generalizability of Pareto-optimization by examining the validity and diversity shrinkage in the Pareto-optimal solutions.	Positive
Porter	2016	[Dissertation] Examined the robustness of Pareto-optimization using selection data from entry-level personnel in the public safety sector of two U.S. municipalities. Results showed that the expected criterion values (i.e., job performance and diversity) obtained from Pareto-optimal weights were consistent with the observed criterion values, supporting the effectiveness of using Pareto-optimization in personnel selection.	Positive
Song, Wee, & Newman	2017	[Journal article] Evaluated the generalizability of Pareto-optimal weights when applied to new samples (i.e., diversity and validity shrinkage). Results showed that diversity and validity shrinkage exist for the Pareto-optimal weighting method. Nonetheless, Pareto-optimization often outperforms unit weighting, despite diversity shrinkage.	Positive
Song, Newman, & Wee	2017	[SIOP symposium] Proposed Pareto-optimal shrinkage formula to adjust for the validity and diversity shrinkage of Pareto-optimal solutions.	Positive
Meade, Thompson, & Schwall	2017	[SIOP symposium] Demonstrated the use of optimization techniques (including ant-colony optimization and Pareto-optimization) in a practical scenario.	Positive
Song, Newman, & Wee	2018	[SIOP symposium] Presented a regularized Pareto-optimal technique, which provides Pareto-optimal solutions with minimal validity and diversity shrinkage when generalized to a new sample.	Positive

(Continued)

Table 2. (Continued)

Authors	Year	Brief summary	Evaluation
Song	2018	[Dissertation] Addressed the Pareto-optimal shrinkage problem by (a) examining Pareto-optimal shrinkage using Monte-Carlo simulations; (b) presenting a formula-based method to adjust for shrinkage in the Pareto-optimal weights; and (c) introducing a regularized-Pareto-optimal algorithm that accounts for shrinkage.	Positive
De Corte, Sackett, & Lievens	2020	[Journal article] A comprehensive simulation study that investigated the performance of the Pareto-optimization method under a large variety of validation selection conditions. Concluded that Pareto-optimal solutions are expected to outperform other weighting methods (e.g., unit weighting).	Positive
<i>Papers that discuss Pareto-optimization</i>			
Sackett & Lievens	2008	[Journal article: Review] In this important review of personnel selection research, the authors discussed Pareto-optimality as a method to reduce adverse impact in exchange for small differences in validity.	Positive
Newman & Lyon	2009	[Journal article: Related study on recruitment] Empirical paper on targeted recruitment strategy. Pareto-optimization was described as an alternative weighting scheme (compared to unit weighting) that can increase diversity with only small decreases in job performance.	Positive
Köhn	2011	[Journal article: Review] Described the mathematical properties underlying Pareto-optimization techniques.	Positive
Sydell, Ferrell, Carpenter, Frost, & Brodbeck	2013	[Book chapter] Described Pareto-optimization as a scoring approach that allows key variables such as diversity and performance to be at the center of attention.	Positive
Oswald, Putka, Ock, Lance, & Vandenberg	2014	[Journal article: Review] Briefly mentioned Pareto-optimality as a way to balance validity and adverse impact concerns.	Positive
Russell, Ford, & Ramsberger	2014	[Technical report] Described Pareto-optimality as a recent approach for weighting different predictors to make personnel selection decisions, and gave a brief, high-level description of how the method can be used.	Positive
Ryan & Ployhart	2014	[Journal article: Review] Referenced Pareto-optimization as an effective means for practitioners to combine selection assessments/predictors and make personnel selection decisions.	Positive
Schmitt	2014	[Journal article: Review] Referenced Pareto-optimization as an analytic technique that allows the pursuit of multiple organizational goals, and described the information needed for the technique. The author mentioned that Pareto-optimization can aid in making decisions on what to include in a selection system.	Positive
Wee, Newman, & Song	2015	[Commentary] Commentary on the use of cognitive subtests in selection. Mentioned Pareto-optimization as a method to minimize the validity/diversity trade-off by choosing appropriate weights for cognitive ability (sub)dimensions.	Positive
Bosco, Allen, & Singh	2015	[Journal article] Referenced Pareto-optimization as a method for selection weighting and mentioned its usefulness for optimizing multiple organizational goals.	Positive
Aiken & Hanges	2017	[Book chapter] Recommended the use of Pareto-optimization in personnel selection to obtain a preferred balance between predicted job performance and diversity in new hires.	Positive
Sackett, Dahlke, Shewach, & Kuncel	2017	[Journal article: Related paper on predictor weighting methods] Suggested that Pareto-optimization could provide gains in predicted job performance and diversity in personnel selection, as compared to the method of unit weighting.	Positive

(Continued)

Table 2. (Continued)

Authors	Year	Brief summary	Evaluation
Darr & Catano	2016	[Journal article: Referred to Pareto-optimization as a future direction] Discussed Pareto-optimization, along with dominance analysis, as methods to better address the trade-offs between validity and adverse impact in selection.	Neutral
Murphy	2018	[Journal article: Review] Discussed the legal context in human resource management and referenced Pareto-optimization as a possible solution to so-called validity-diversity trade-offs.	Positive
Rickes	2018	[Dissertation] Discussed Pareto-optimization as a potential method to improve the prediction of undergraduate first-year graduate GPA and reduce adverse impact.	Positive

Overall, despite the absence of case law on Pareto-optimal weighting methods, our examination of the extent to which the method meets the Daubert standards demonstrates that the method has a reasonable likelihood of being considered both legally and scientifically valid.

Practical recommendations

We close with the presentation of Table 3, which provides some additional practical tips to keep in mind when applying Pareto-optimization. This includes carefully considering how job performance is operationalized, the samples used for calibration and pilot testing, the timing of data collection, and how selection decisions will be made once optimized composite scores are computed. It also contains a reminder to users of the benefit of collecting content validity evidence for all predictors, and to not get so caught up in the “metrics” that careful, qualitative consideration of item-level content is neglected. Finally, we recommend systematic and continuous legal auditing, in ways that protect the organization from legal scrutiny as it seeks to proactively improve the performance and diversity of its workforce. Here we highlight some of the more complex issues inherent to implementing Pareto-optimal weights within personnel selection.

Individual predictors are still vulnerable to legal challenge

Throughout this article, we have provided evidence that the Pareto-optimization method shows promise for reducing adverse impact, maintaining job performance prediction, and withstanding legal scrutiny. That being said, organizations may still be vulnerable to adverse impact claims made about individual predictors within their selection systems. Specifically, in *Connecticut v. Teal* (1982), the Supreme Court ruled that a lack of adverse impact in the total selection system (i.e., the bottom-line defense) does not preclude plaintiffs from successfully showing disparate impact against individual components of the selection system. Therefore, each selection predictor utilized within a selection system must demonstrate sufficient, individual evidence of validity (i.e., job relatedness as specified in the *Uniform Guidelines* [EEOC, 1978]), and alternative predictors should not exist that measure the same knowledge, skills, abilities, and other characteristics (KSAOs) with similar predictive power but less adverse impact (cf. *Griggs v. Duke Power Co.*, 1971).

Choice of performance criterion matters

As we have highlighted throughout this article, there are a number of issues to consider pertinent to the measurement of job performance. Although our case example employed a unidimensional measure of task performance as the performance criterion, a wider conceptualization of job

Table 3. A checklist of the key decisions for adopting Pareto-optimization for personnel selection

Stage 1. Preparation
<p>a) Determine goals of selection: choose the criteria to optimize</p> <ul style="list-style-type: none"> <input type="checkbox"/> What are the main criteria of interest? (e.g., job performance, diversity). <input type="checkbox"/> What are the specific subgroups of interest? <input type="checkbox"/> Decide the extent to which each outcome will be optimized (e.g., some organizations may choose to select solutions that allow more favorable adverse impact ratios only to the extent that performance validity is not meaningfully decreased). <input type="checkbox"/> Operationalize the criterion. Decision makers should carefully consider the measurement of job performance, considering factors such as type of performance (e.g., task, contextual, maximum, typical); interrater unreliability; training and structure of assessments to minimize error in measurement; range restriction in performance scores; and timing of measurement given the dynamic nature of performance (e.g., before or after beginner learning curve). Further, decision makers must decide how they will input the performance criterion into the Pareto-optimization model (e.g., job performance criterion validity, expected average job performance, selection utility). <input type="checkbox"/> Evaluate the content of the performance criterion (including individual items). <input type="checkbox"/> Job analysis should be used as the basis for decisions. <p>b) Choose predictors and assessment methods</p> <ul style="list-style-type: none"> <input type="checkbox"/> What predictors best predict the criteria of interest? <input type="checkbox"/> How to assess the predictors? Are the assessment tools already available or do they need to be developed? (e.g., work-sample tests, situational judgment tests, structured interview, etc.). <input type="checkbox"/> Evaluate the content of the predictor (including individual items). <input type="checkbox"/> Each predictor should individually demonstrate evidence of job-relatedness (<i>Connecticut v. Teal</i>, 1982). The <i>Uniform Guidelines</i> (EEOC, 1978) specifies that either content or criterion-related validity evidence can be used, depending on the appropriateness of the context. Ideally, each predictor would show evidence of both. <input type="checkbox"/> Will the selection process follow a compensatory and top-down procedure or a multiple hurdles or multi-stage procedure (De Corte, 2014; De Corte et al., 2011)? <p>c) Obtain the Pareto-optimal predictor weights</p> <ul style="list-style-type: none"> <input type="checkbox"/> Calibration sample: What employee group to sample as the calibration sample? What is the targeted sample size? Incumbent sample should match the expected applicant pool as closely as possible. The calibration sample size should be as large as possible (ideally greater than 100; cf. Song, Wee, & Newman, 2017). <input type="checkbox"/> <u>Table 1. Step 1.</u> Collect data from <i>current employees</i>. <input type="checkbox"/> <u>Table 1. Step 2.</u> Compute input statistics for Pareto-optimization. <input type="checkbox"/> <u>Table 1. Step 3.</u> Obtain the Pareto-optimal predictor weights using statistics from Table 1 Step 2. <input type="checkbox"/> Be sure to consider any time lags between, and potential reverse order of, the collection of incumbent predictor and criterion (performance) data. When appropriate, prepare for additional, subsequent criterion data collection, and make data retention and management plans. <input type="checkbox"/> Carefully document the above procedures (e.g., data collection and storage, estimation, decision making). <p>d) Pilot trial (highly recommend)*</p> <ul style="list-style-type: none"> <input type="checkbox"/> <u>Table 1. Step 3c (pilot trial).</u> Test the Pareto-optimal predictor weights in the selection of job <i>applicants</i>. The purpose of the pilot trial is to test Pareto-optimization in the selection scenario. Thus, in this circumstance, the method is not used for making the actual selection decisions. <input type="checkbox"/> Estimate the criterion in the validation sample (e.g., validation sample job performance criterion validity). <input type="checkbox"/> Evaluate shrinkage using the applicant (validation) sample for both job performance criterion validity and the adverse impact ratio [e.g., validity shrinkage: comparison of the (expected) job performance criterion validity of the calibration sample vs. the job criterion (actual) validity of the validation sample; see Song, Wee, & Newman, 2017 for details]. The closer the match between the expected and actual job performance validity values, the better evidence that the Pareto-optimization solution is useful for future applicant samples. In the example provided herein, this would mean that if predictor 1, 2, and 3 are weighted at .01, .22, and .76, respectively, then we would want to examine how close the performance validity is to .36 and how close the AI ratio is to .82 in the applicant sample.
Stage 2. Implementation
<p>a) Use Pareto-weighted predictor composite to make selection decisions</p> <ul style="list-style-type: none"> <input type="checkbox"/> <u>Table 1. Step 4.</u> Implement the Pareto-optimal predictor weights in the selection of job <i>applicants</i>.
Stage 3. Legal audit
<ul style="list-style-type: none"> <input type="checkbox"/> Engage legal counsel to examine the legal intricacies associated with the use of this method under specific selection contexts and to navigate the complexities of the coverage of data, tests, and analyses under privilege (e.g., Kaufmann, 2009; Walkowiak, 2008). <input type="checkbox"/> Carefully document the selection process. <input type="checkbox"/> Do not use current applicants to determine Pareto-optimization weights. As discussed above, weights should be established using a calibration sample <i>a priori</i>.

(Continued)

Table 3. (Continued)

Stage 4. Maintenance
<input type="checkbox"/> Evaluate selection criterion (e.g., evaluate the adverse impact ratio of multiple subgroups) in each round of selection.
<input type="checkbox"/> Evaluate shrinkage (e.g., validity and diversity shrinkage, see Song, Wee, & Newman, 2017).
<input type="checkbox"/> Monitor change among applicant samples (e.g., across seasonal recruitment cycles).

*If a pilot trial is not practically feasible, see Song, Wee, & Newman (2017) for computational methods to evaluate the Pareto-optimal predictor weights when applied to the applicant sample.

performance allows for a wider array of predictors, including those that are more resistant to adverse impact (e.g., personality measures). This is relevant for Pareto-optimization (as well as other predictor-weighting methods), in that a wider “criterion space” provides more latitude for estimating predictor weighting schemes that maximize the prediction of performance and minimize adverse impact. Research on the use of multiple performance criteria within Pareto-optimization is still developing. This work has combined multiple performance criteria (e.g., task performance, contextual performance) into a weighted composite using prespecified weights. For example, De Corte et al. (2007) created a weighted sum of standardized task performance and contextual performance scores, with task performance weighted three times that of contextual performance. The authors chose the 1:3 criterion weights based on past research and subsequently examined the performance outcome of the Pareto-optimal solutions as the mean weighted-performance score obtained by the selected candidates. They found that there was still a relatively steep trade-off between performance and diversity criteria, even when contextual performance was included as a part of overall performance. Decisions on the types of performance to include and the computation of composite performance measures should be based on the specific organization and job context. Importantly, all job performance measures should be preceded by and based on job analysis (EEOC, 1978; Gatewood et al., 2015).

A second set of performance-related issues to consider pertains to the psychometric quality of the performance measures used to calculate the criterion-related validities that are input into the Pareto-optimization algorithm. As performance data often take the form of supervisor ratings, the reliability of the resulting performance scores often suffers due to lack of or insufficient rater training and clear rubrics with which to make ratings. Thus, in order to obtain accurate validity estimates as input into the Pareto-optimization algorithm, improvements might need to be made to the performance measurement system (again, based on job analysis). Further, as these data are collected on incumbents rather than applicants, performance (as well as predictor) scores are generally range-restricted. Thus, the predictor intercorrelations and the criterion-related validity of each predictor need to be corrected for range restriction (De Corte, 2014; De Corte et al., 2011; Roth et al., 2011).

Finally, we should note that although our case example used job performance criterion validity as input into the Pareto algorithm, there are other computations that could be input here instead. One alternative is expected average job performance (also referred to as the “expected average criterion score” and the “average job performance of the selected employees”), which is the average of the expected job performance score of the selected applicants (see De Corte, 2014; De Corte et al., 2007, 2020; Druart & De Corte, 2012a, 2012b). Another alternative is selection utility, which refers to the job performance gain of Pareto-optimization over random selection, after taking into account costs (for details, see the appendix in De Corte et al., 2007).

Multiple hurdle (and other) selection strategies

Pareto-optimization is most often discussed as a weighting technique to be applied within a compensatory, top-down selection process. However, it is also applicable to multiple hurdle or

multistage selection designs (see De Corte *et al.*, 2011). The analytical procedures for multistage Pareto-optimization are generally similar to the single-stage scenario, with the exception of the analytical expression of selection outcomes (e.g., job performance and diversity), which considers the noncompensatory characteristics of the predictors (see De Corte *et al.*, 2006, for details). De Corte (2014) provides a computational tool, COPOSS, to implement Pareto-optimization within a multistage scenario, while De Corte *et al.* (2011) provide a computational software, SSOV, as a decision aid for designing Pareto-optimal selection systems (including the multistage selection scenario). Both tools as well as their tutorials are available for download at <https://users.ugent.be/~wdecorte/software.html>. Compared to the single-stage setting, multistage optimization is more cost-effective, but will likely yield less optimal validity/diversity trade-off solutions. This is because multistage optimization usually can only use part of the predictor information at a time, compared to single-stage optimization, which makes use of all available predictors simultaneously. However, given the complexity of selection systems (including predictor characteristics, applicant pool distribution, contextual constraints), the superiority of a single- vs. multistage strategy depends on the context.

Judgment calls still required, especially when multiple subgroup comparisons come into play

As we have noted, although Pareto-optimization provides a systematic and quantitative approach to selecting predictor weights that seek to maximize job performance prediction and minimize adverse impact, the method still requires judgment calls at multiple stages of its implementation. The organization must carefully consider not only how to measure performance and which predictors to use, but also how validity and diversity will be prioritized within the selection system.

Further, it must decide which subgroup contrasts are relevant for consideration, and it must face the reality that different (perhaps opposing) weighting solutions may be “optimal” for different contrasts (e.g., the solution that maximizes validity and minimizes Black/White subgroup differences may differ from that which minimizes subgroup differences between men and women). To address this issue, Song and Tang (2020) developed an updated Pareto-optimal technique to simultaneously consider multiple subgroup comparisons, which includes multiple subgroups within the same demographic category (e.g., race), as well as multiple demographic categories (e.g., race and gender). The magnitude of the validity-diversity tradeoff in multi-subgroup optimization is influenced by (1) the subgroup mean differences between the majority group and minority groups and (2) the subgroup mean differences among the minority groups. Using Monte-Carlo simulation, Pareto-optimal weighting for multiple subgroups is currently being evaluated in various selection scenarios (e.g., where the proportion of minority applicants varies). This research will provide guidance on how likely “opposing” Pareto-optimal solutions among different subgroup contrasts actually are, and ways in which Pareto-optimal weighting schemes, which consider multiple comparisons simultaneously, could be obtained.

Conclusion

In this practice forum, we sought to highlight the tensions organizations face when seeking to create and implement effective, fair, and legally compliant personnel selection systems. We traced the history of innovations presented by I-O psychologists in reconciling the so-called diversity-validity tradeoff and presented Pareto-optimization as a potential way forward to systematically optimize performance and diversity criteria. We then provided a primer to the method at varying levels of sophistication and presented user-friendly tools for implementing the technique in practice. Finally, we attempted to scrutinize the method from an EEO law perspective, and in doing so offered potential defenses that might be waged in justifying the method if challenged.

It is important to note that discussion of Pareto-optimization is generally limited to academic outlets. Our search revealed no case analyses describing the method as used in practice, nor legal

analysis such as that provided here. We hope this article will encourage practitioners to submit examples to this forum to better highlight the benefits and challenges associated with the application of Pareto-optimization, and for those in the legal arena to weigh in with their views on the legal appropriateness of the method.

References

- Aiken, J. R., & Hanges, P. J. (2017). The sum of its parts: Methods of combining assessments for employment decisions. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 388–396). New York, NY: Routledge.
- Black Law Enforcement Officers Assoc. v. Akron*, U.S. Dist. LEXIS 30160 (1986).
- Bobko, P. (2001). *Correlation and regression: Applications for industrial organizational psychology and management*. Thousand Oaks, CA: Sage.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black–White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology*, *66*, 91–126.
- Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, *10*, 689–709.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52*, 561–589.
- Bosco, F., Allen, D. G., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology*, *68*, 859–898.
- Bureau of Labor Statistics. (2000). *Employment and earnings*, 46(1), 13–14 (Chart A-4).
- Campion, M. A., Outtz, J. L., Zedeck, S., Schmidt, F. L., Kehoe, J. F., Murphy, K. R., & Guion, R. M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology*, *54*, 149–185.
- Cascio, W., Jacobs, R., & Silva, J. (2010). Validity, utility, and adverse impact: Practical implications from 30 years of data. In J. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 217–288). New York: Routledge.
- Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management*. Harlow, Essex, UK: Pearson Education Limited.
- Civil Rights Act of 1964 § 7, 42 U.S.C. § 2000e et seq (1964).
- Connecticut v. Teal*, 457 US 440. (1982).
- Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the black–white gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology*, *100*, 1713–1736.
- Darr, W. A., & Catano, V. M. (2016). Determining predictor weights in military selection: An application of dominance analysis. *Military Psychology*, *28*, 193–208.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
- Das, I., & Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, *8*(3), 631–657.
- De Corte, W. (2014). Predicting the outcomes of single- and multistage selections when applicant pools are small and heterogeneous. *Organizational Research Methods*, *17*, 412–432.
- 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*(3), 260–270.
- De Corte, W., Sackett, P. R., & Lievens, F. (2011). Designing Pareto-optimal selection systems: Formalizing the decisions required for selection system development. *Journal of Applied Psychology*, *96*, 907–926.
- De Corte, W., Sackett, P. R., & Lievens, F. (2020). Robustness, sensitivity, and sampling variability of Pareto-optimal selection system solutions to address the quality-diversity trade-off. *Organizational Research Methods*, *23*, 535–568. <https://doi.org/10.1177/1094428118825301>
- 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.
- De Soete, B., Lievens, F., & Druart, C. (2013). Strategies for dealing with the diversity-validity dilemma in personnel selection: Where are we and where should we go? *Journal of Work and Organizational Psychology*, *29*(1), 3–12.
- Druart, C., & De Corte, W. (2010, April). Optimizing the efficiency: Adverse impact trade-off in personnel classification decisions. Paper presented at the 25th Annual Conference of the Society for Industrial and Organizational Psychology.

- Druart, C., & De Corte, W.** (2012a). Designing Pareto-optimal systems for complex selection decisions. *Organizational Research Methods*, *15*, 488–513.
- Druart, C., & De Corte, W.** (2012b). Computing Pareto-optimal predictor composites for complex selection decisions. *International Journal of Selection and Assessment*, *20*, 385–393.
- Einhorn, H. J., & Hogarth, R. M.** (1975). Unit weighting schemes for decision making. *Organizational Behavior & Human Performance*, *13*, 171–192.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor & Department of Justice (EEOC).** (1978). *Uniform guidelines on employee selection procedures*. *Federal Register*, *43*, 38290–39315.
- Finch, D. M., Edwards, B. D., & Wallace, J. C.** (2009). Multistage selection strategies: Simulating the effects on adverse impact and expected performance for various predictor combinations. *Journal of Applied Psychology*, *94*, 318–340.
- Gatewood, R., Feild, H. S., & Barrick, M.** (2015). *Human resource selection*. Boston, MA: Cengage Learning.
- Griggs v. Duke Power Co.**, 401 U.S. 424 (1971).
- Gutman, A., Koppes, L. L., & Vodanovich, S. J.** (2011). *EEO law and personnel practices*. New York, NY: Psychology Press.
- Gutman, A., Outtz, J. L., Dunleavy, E.** (2017). An updated sampler of legal principles in employment selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 631–658). New York, NY: Routledge.
- 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.
- Hayden v. County of Nassau**, 180 F.3d 42 (2nd Cir. 1999).
- Henle, C. A.** (2004). Case review of the legal status of banding. *Human Performance*, *17*, 415–432.
- Kaufmann, P. M.** (2009). Protecting raw data and psychological tests from wrongful disclosure: A primer on the law and other persuasive strategies. *The Clinical Neuropsychologist*, *23*, 1130–1159.
- Kaye, D. H.** (2001). The dynamics of Daubert: Methodology, conclusions, and fit in statistical and econometric studies. *Virginia Law Review*, *87*, 1933–2018.
- Kehoe, J. F.** (2008). Commentary on Pareto-optimality as a rationale for adverse impact reduction: What would organizations do? *International Journal of Selection and Assessment*, *16*, 195–200.
- Köhn, H. F.** (2011). A review of multiobjective programming and its application in quantitative psychology. *Journal of Mathematical Psychology*, *55*, 386–396.
- Kraft, D.** (1988). A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt*. Retrieved from http://degenerateconic.com/wp-content/uploads/2018/03/DFVLR_FB_88_28.pdf.
- Lievens, F.** (2015). Diversity in medical school admission: Insights from personnel recruitment and selection. *Medical Education*, *49*, 11–14.
- Lopes, L. G. D., Brito, T. G., Paiva, A. P., Peruchi, R. S., & Balestrassi, P. P.** (2016). Robust parameter optimization based on multivariate normal boundary intersection. *Computers & Industrial Engineering*, *93*, 55–66.
- McKay, P. F., & McDaniel, M. A.** (2006). A reexamination of black-white mean differences in work performance: More data, more moderators. *Journal of Applied Psychology*, *91*, 538–554.
- Meade, A. W., Thompson, I. B., & Schwall, A. R.** (2017). Optimizing validity while controlling adverse impact with ant colony optimization. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Morris, S.** (2017). Statistical significance testing in adverse impact analysis. In S. B. Morris & E. M. Dunleavy (Eds.), *Adverse impact analysis: Understanding data, statistics and risk* (pp. 92–112). New York, NY: Routledge.
- Motowidlo, S. J., & Kell, H. J.** (2012). Job performance. In I. Weiner, N. W. Schmitt, & S. Highhouse (Eds.), *Handbook of psychology*, Vol. 12: *Industrial and organizational psychology* (2nd ed., pp. 82–103). Hoboken, NJ: John Wiley & Sons, Inc.
- Murphy, K. R.** (2018). The legal context of the management of human resources. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*, 157–182.
- Naves, F. L., de Paula, T. I., Balestrassi, P. P., Braga, W. L. M., Sawhney, R. S., & de Paiva, A. P.** (2017). Multivariate normal boundary intersection based on rotated factor scores: A multiobjective optimization method for methyl orange treatment. *Journal of Cleaner Production*, *143*, 413–439.
- Newman, D. A., Jacobs, R. R., & Bartram, D.** (2007). Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal of Applied Psychology*, *92*(5), 1394–1413.
- Newman, D. A., Jones, K. S., Fraley, R. C., Lyon, J. S., & Mullaney, K. M.** (2014). Why minority recruiting doesn't often work, and what can be done about it: Applicant qualifications and the 4-group model of targeted recruiting. In K. Y. T. Yu & D. M. Cable (Eds.), *The Oxford handbook of recruitment* (pp. 492–526). New York, NY: Oxford University Press.
- Newman, D. A., & Lyon, J. S.** (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology*, *94*, 298–317.
- Nimmegeers, P., Telen, D., Logist, F., & Van Impe, J.** (2016). Dynamic optimization of biological networks under parametric uncertainty. *BMC Systems Biology*, *10*(1), 86.

- Oswald, F. L., Dunleavy, E., & Shaw, A. (2017). Measuring practical significance in adverse impact analysis. In S. B. Morris & E. M. Dunleavy (Eds.), *Adverse impact analysis: Understanding data, statistics and risk* (pp. 92–112). New York, NY: Routledge.
- Oswald, F. L., Putka, D. J., Ock, J., Lance, C. E., & Vandenberg, R. J. (2014). Weight a minute, what you see in a weighted composite is probably not what you get. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 187–205). London, UK: Routledge.
- Oultz, J. L., & Newman, D. A. (2010). A theory of adverse impact. In J. L. Oultz (ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 80–121). New York: Routledge.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, **61**, 153–172.
- Porter, M. G. (2016). *An application of Pareto-optimality to public safety selection data: Assessing the feasibility of optimal composite weighting* (Doctoral dissertation, Illinois Institute of Technology, Chicago, Illinois). Retrieved from <https://search.proquest.com/dissertations/docview/1802060511/fulltextPDF/BE33EB0CF10C486APQ/1?accountid=13360>
- Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment*, **13**(4), 304–315.
- Potosky, D., Bobko, P., & Roth, P. L. (2008). Some comments on Pareto thinking, test validity, and adverse impact: When ‘and’ is optimal and ‘or’ is a trade-off. *International Journal of Selection and Assessment*, **16**, 201–205.
- Pyburn Jr., K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity–validity dilemma: Overview and legal context. *Personnel Psychology*, **61**, 143–151.
- Reynolds v. Ala. DOT*, 295 F. Supp. 2d 1298 (2003).
- Ricci v. DeStefano*, 557 U.S. 557 (2009).
- Rickes, S. (2018). *Traditional and nontraditional predictors of college FYGPA and racial/ethnic subgroup differences*. (Doctoral dissertation, Alliant International University, San Diego, CA).
- Roth, P. L., BeVier, C. A., Bobko, P., Switzer III, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, **54**, 297–330.
- Roth, P. L., Switzer III, F. S., Van Iddekinge, C. H., & Oh, I. S. (2011). Toward better meta-analytic matrices: How input values can affect research conclusions in human resource management simulations. *Personnel Psychology*, **64**, 899–935.
- Russell, T. L., Ford, L., & Ramsberger, P. (2014). *Thoughts on the future of military enlisted selection and classification* (HumRRO Technical Report No. 053). Alexandria, VA: Human Resources Research Organization (HumRRO).
- Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, **65**, 693–717.
- Sackett, P. R., Dahlke, J. A., Shewach, O. R., & Kuncel, N. R. (2017). Effects of predictor weighting methods on incremental validity. *Journal of Applied Psychology*, **102**(10), 1421–1434.
- Sackett, P. R., De Corte, W., & Lievens, F. (2008). Pareto-optimal predictor composite formation: A complementary approach to alleviating the selection quality/adverse impact dilemma. *International Journal of Selection and Assessment*, **16**, 206–209.
- Sackett, P. R., De Corte, W., & Lievens, F. (2010). Decision aids for addressing the validity-adverse impact tradeoff: Implications for organizational staffing and high stakes selection. In J. Oultz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 459–478). New York: Routledge.
- Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, **50**, 707–721.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, **59**, 419–450.
- Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review Organizational Psychology and Organizational Behavior*, **1**, 45–65.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.
- Society for Industrial and Organizational Psychology (SIOP). (2018). Principles for the validation and use of personnel selection procedures (5th ed.). *Industrial and Organizational Psychology*, **11**(S1), 1–97.
- Song, Q. C. (2018). *Diversity shrinkage of Pareto-optimal solutions in hiring practice: Simulation, shrinkage formula, and a regularization technique* (Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL).
- Song, Q. C., Newman, D. A., & Wee, S. (2017, June). Approximation of diversity shrinkage from Pareto weights for diversity-performance tradeoffs. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Song, Q. C., Newman, D. A., & Wee, S. (2018, April). Enhancing diversity: Pareto-optimal weighting algorithm with regularization. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.
- Song, Q. C., & Tang, C. (2020, June). Adverse impact reduction for multiple subgroups: A Pareto-optimization approach. In Q. C. Song (Co-Chair) and S. Wee (Co-Chair), *Multi-Objective Optimization in the Workplace: Addressing Adverse Impact in Selection*. Symposium presentation at the 35th Annual Conference of the Society for Industrial and Organizational Psychology, Virtual Conference.

- Song, Q. C., Wee, S., & Newman, D. A.** (2016, April). Cross-validating Pareto-optimal weights for reducing adverse impact. In P. J. Hanges & J. Y. Park (Co-chairs), *New Insights into Adverse Impact: Origination, Motivation, and Scale Weighting*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.
- Song, Q. C., Wee, S., & Newman, D. A.** (2017). Diversity shrinkage: Cross-validating Pareto-optimal weights to enhance diversity via hiring practices. *Journal of Applied Psychology*, *102*, 1636–1657.
- Sydell, E., Ferrell, J., Carpenter, J., Frost, C., & Brodbeck, C. C.** (2013). Simulation scoring. In M. Fetzter & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 83–107). New York, NY: Springer.
- Tam, A. P., Murphy, K. R., & Lyall, J. T.** (2004). Can changes in differential dropout rates reduce adverse impact? A computer simulation study of a multi-wave selection system. *Personnel Psychology*, *57*, 905–934.
- Tsang, H.** (2010). *Improving the adverse impact and validity trade-off in Pareto optimal composites: A comparison of weights developed on contextual vs task performance differences* (Doctoral dissertation, University of Central Florida, Orlando, FL).
- Uniform Guidelines on Employee Selection Procedure** (1978); 43 FR ____ (August 25, 1978).
- Walkowiak, V. S.** (2008). An overview of the attorney-client privilege when the client is a corporation. In V. S. Walkowiak (ed.), *The attorney-client privilege in civil litigation: Protecting and defending confidentiality* (pp. 1–56). Chicago, IL: American Bar Association.
- Wee, S., Newman, D. A., & Joseph, D. L.** (2014). More than *g*: Selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology*, *99*, 547–563.
- Wee, S., Newman, D. A., & Song, Q. C.** (2015). More than *g*-factors: Second-stratum factors should not be ignored. *Industrial and Organizational Psychology*, *8*, 482–488.

APPENDIX. Pareto optimization: Technical details

Pareto-optimal weighting for multi-objective optimization (to create Pareto curves; e.g., Figure 1) can be implemented in a variety of ways, one of which is labeled *normal boundary intersection* (NBI; see Das & Dennis [1998] for a foundational introduction to this method). The aim of the algorithm is to find evenly spaced sets of solutions on the Pareto curve that optimize multiple criteria (e.g., diversity and job performance) under certain constraints (e.g., the variance of the composite predictor is 1 [in order to find unique solutions]).

An example of NBI is shown in Figure A1. The horizontal axis shows the demographic diversity of new hires (represented by adverse impact ratio), whereas the vertical axis shows their expected job performance (represented by the job performance validity of the predictor composite). The blue oval (which includes both the solid and dotted blue border) represents the solution space of all possible solutions under certain constraints.

There are three main steps in the NBI algorithm.

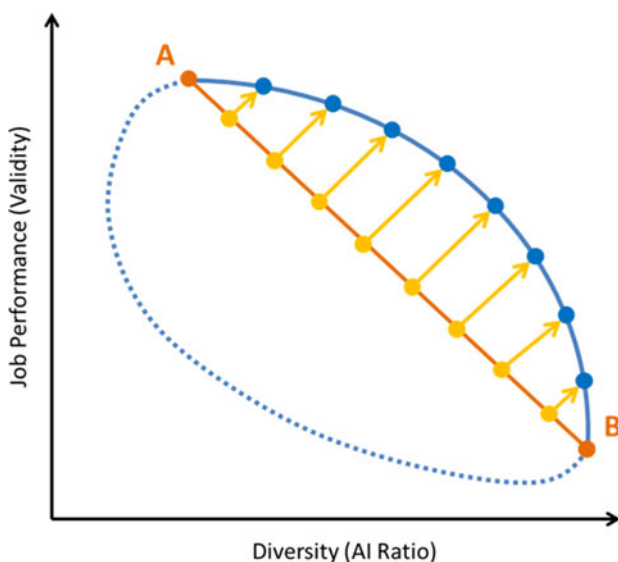


Figure A1. An illustration of the normal boundary intersection (NBI) algorithm.

Step 1: Find the endpoints (e.g., Points A and B in Figure 1) and corresponding predictor weights. Specifically, the SLSQP algorithm¹⁸ is used to find one set of predictor weights (Point A) where only job performance is maximized; and another set of predictor weights (Point B) where only diversity (represented using adverse impact ratio) is maximized.

Step 2: Linear interpolation of evenly-spaced solutions between the endpoints. To find the Pareto points between the two endpoints (found in Step 1), the algorithm first creates a line that connects the two endpoints (i.e., the orange line in Figure A1) and specifies evenly spaced points along this line. The number of points along the line (i.e., including the two end points; yellow dots in Figure A1) equals the user-specified number of Pareto solutions (e.g., 21 evenly spaced solutions).

Step 3: Projection of evenly spaced solutions between the endpoints. At each Pareto point, the SLSQP algorithm is again used to find the optimal set of weights. Specifically, the algorithm will project in a perpendicular direction from the line created in Step 2 (i.e., yellow arrows in Figure A1) until it reaches the border of the solution space (i.e., blue oval in Figure A1), finding a set of Pareto-optimal predictor weights (i.e., a blue dot in Figure A1). This process will iterate through all Pareto points (e.g., 21 points) until the optimal predictor weights for each Pareto point are obtained.

¹⁸Sequential least squares programming (SLSQP) is an algorithm that finds a set of weights that maximizes the criteria, given the constraints. It is a least-squares method similar to that used in least squares regression. See Kraft (1998) for more information.

Cite this article: Rupp, D.E., Song, Q.C., and Strah, N. (2020). Addressing the so-called validity–diversity trade-off: Exploring the practicalities and legal defensibility of Pareto-optimization for reducing adverse impact within personnel selection. *Industrial and Organizational Psychology* 13, 246–271. <https://doi.org/10.1017/iop.2020.19>