Journal of Applied Psychology
2011, Vol. 96, No. 5, 907–926

0021-9010/11/$12.00 DOI: 10.1037/a0023298

Designing Pareto-Optimal Selection Systems: Formalizing the Decisions Required for Selection System Development

Wilfried De Corte
Ghent University

Paul R. Sackett
University of Minnesota, Twin Cities Campus

Filip Lievens
Ghent University

The article presents an analytic method for designing Pareto-optimal selection systems where the applicants belong to a mixture of candidate populations. The method is useful in both applied and research settings. In an applied context, the present method is the first to assist the selection practitioner when deciding on 6 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. From a research perspective, the method offers a unique opportunity for studying the impact and relative importance of different strategies for reducing adverse impact.

Keywords: adverse impact, personnel selection, Pareto-optimal, selection design

This article presents an analytic method for designing Pareto-optimal selection systems where the applicants belong to a mixture of candidate populations. To set the stage, consider the situation where a battery of predictors is available for selecting a given number of employees from an applicant pool that is a mixture of majority and one or more minority candidate groups. When both the goals of selecting a high-quality and a diverse work force are valued, the selection practitioner faces several decisions. First is the decision as to how many and which of the available predictors to use given prevailing constraints with regard to, for example, testing time and overall selection costs. This is the *predictor subset decision* problem. The next decision, henceforth referred to as the *selection rule decision*, involves choosing between a compensatory and a non-compensatory selection scheme (or a combination of both). In a compensatory scheme, lower scores on one predictor may be compensated for by higher scores on other predictors, whereas the decision to maintain a given minimum cutoff level for at least one of the predictors leads to a non-compensatory scheme. A third decision issue concerns the choice between single and multiple stage designs: the *selection staging decision*. In a single-stage scheme, all the predictors are administered prior to any screening decisions, whereas in multi-stage schemes, the predictors are administered in several stages, with only those that pass the previous stage(s) moving on to subsequent stages. Observe that we distinguish between multi-stage and multiple hurdle selection schemes because a non-compensatory single-stage scheme also amounts to a multiple hurdle selection in that all of the candidates are in that case first screened on all of the non-compensatory predictors with the final selection done among those passing all of the hurdles.

When opting for a multi-stage approach, the *sequencing of the predictors* over the stages constitutes a fourth decision in the design process. Together, the predictor subset decision, the selection rule decision, and the selection staging decision, eventually completed with the predictor sequencing decision, lead to the chosen *selection scenario*. The fifth decision concerns the *weighting of the predictors*. The predictor weighting issue occurs as soon as several predictors are combined into either the single-stage predictor composite or any one of the multi-stage composites in multi-stage scenarios. Sixth, and finally, but only in case of multi-stage selection decisions, *retention decisions* must be made, specifying at each intermediate stage the proportion of initial applicants who will be retained for further scrutiny in the next stage.

In general, the above listed decisions are mutually dependent implying that when designing selections, one should consider all six decisions simultaneously. The method we present here is the first, to our knowledge, to provide such an integrated approach by guiding the selection researcher and practitioner through each of the six decision stages. The guidance results in selection designs that show a Pareto-optimal trade-off between the goals pursued by the selection. The method integrates earlier work on (a) the estimation of selection outcomes in multi-stage selection (De Corte, Lievens, & Sackett, 2006), (b) the formation of Pareto-optimal predictor composites in single-stage selection (De Corte, Lievens, & Sackett, 2007), and (c) the selection of optimal predictor subsets (De Corte, Sackett, & Lievens, 2010; Johnson, Abrahams, & Held, 2004) within an overarching approach to optimal selection design. The method is useful in both applied and research settings. As an applied tool, the method may assist the selection practitioner in shaping selections that lead to optimal trade-offs between valued

This article was published Online First April 4, 2011.

Wilfried De Corte, Department of Data Analysis, Ghent University, Ghent, Belgium; Paul R. Sackett, Department of Psychology, University of Minnesota, Twin Cities Campus; Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University.

Correspondence concerning this article should be addressed to Wilfried De Corte, Department of Data Analysis, Ghent University, H. Dunantlaan 1, 9000 Ghent, Belgium. E-mail: wilfried.decorte@ugent.be

selection outcomes, whereas the selection researcher can use the method to study, for example, the impact and relative importance of several different strategies to reduce adverse impact (cf. Ployhart & Holtz, 2008).

We refer to a *selection system* as the operational selection design that results from a particular set of choices with respect to the six major selection decision areas. To illustrate this, consider the situation where one plans selecting 10% of the applicants and where the set of available predictors consists of (a) a cognitive ability (CA) test, (b) a structured interview (SI), (c) a conscientiousness (CO) measure, (d) a biodata (BI) questionnaire, and (e) an integrity (IN) test. In that case, deciding to use only the first four predictors (predictor subset decision) with a minimum cutoff value for CA (selection rule decision), within a two-stage selection setup (selection staging decision) in which CA, CO, and BI are administrated in the first stage and the SI in the final stage (predictor sequencing decision), and where the first-stage retention decision is based on a regression weighted composite of CO and BI (predictor weighting decision), retaining 25% of the initial candidates after the first stage (retention decision), exemplifies a selection system. Note that this selection system, although a two-stage system, effectively involves three hurdles because an accepted candidate must (a) score at or above the CA cutoff, (b) score sufficiently high on the CO and BI composite to be retained for the second-stage predictor administration, and (c) belong to the 40% (i.e., 10/25 $\times$ 100) highest scoring candidates on the SI predictor in the candidate group remaining after the first stage. Also, observe the difference between the cutoff involved in the first hurdle and the cutoffs implied at the end of the first stage (second hurdle) and the final stage (third hurdle). The first cutoff is given explicitly as a required minimum value on the CA predictor, whereas the second and third hurdle cutoffs are implied by the chosen retention (25% after Stage 1) and the overall selection rate (10%), respectively.

Given the many decisions required and the particularities of the selection situation at hand, many different selection systems will typically be feasible, but some of these systems are likely to result in a combination of selection outcome values (e.g., expected average job performance of the selected candidates or adverse impact ratio [AIR]) that cannot be bettered by any other feasible selection system: These systems show a Pareto-optimal trade-off between the valued goals pursued in the selection. More specifically, a selection system is called *Pareto-optimal* (or, equivalently, shows a Pareto-optimal trade-off) when no other feasible selection system results in outcome values that are all at least as favorable (and at least one outcome is more favorable) than the outcomes associated with the first system.

To arrive at the set of Pareto-optimal selection systems, the present method will invoke certain assumptions (cf. Appendix A); use limited information on the makeup of the initial applicant pool; and require data on the validity, effect size (i.e., subgroup difference), and intercorrelation of the predictors (cf. Appendix B). Alternatively, the method has the flexibility to adapt to a diversity of contextual features. For example, it will be shown how the method can accommodate contextual constraints on, among others, the total selection costs and time, the maximum number of administered predictors, and the sequencing of predictors.

The next section describes the new method for designing Pareto-optimal selection systems. The description goes hand in hand with an example application that illustrates the practical contribution of the method. Because the method is intended as a decision aid in the design of future selection systems, it will typically be applied using estimates rather than the true predictor and applicant data values. We therefore subsequently investigate whether the results of the method, as derived from current best estimates of the predictor and applicant input data, generalize to situations where the actually obtained data values show realistic sampling variability relative to the initial estimated data. Next, the method is used to study the impact of various methods of weighting predictors (e.g., regression weights vs. unit weights) when creating predictor composites. The relative merits of compensatory versus non-compensatory and single-stage versus multi-stage designs in generating favorable selection quality/work force diversity trade-offs are also investigated. A final research contribution addresses the common practice in compensatory selection of using composites of only the stage-specific administered predictors instead of composites from these as well as previously administered predictors.

Obtaining Pareto-Optimal Selection Systems

The design of Pareto-optimal selection systems consists of two steps. The first, preparatory step typically includes the following three subtasks: (1) choice and articulation of the selection goals, (2) inventory and characterization of the available predictor battery and the current or expected applicant pool, and (3) specification of the relevant contextual constraints such as, for example, prevailing test cost considerations and/or practical and logistical limitations on the sequencing of the predictors. The results of the first step are then used in the second step to define the parameters and boundary conditions of the optimization problem that, when resolved by applying the present analytic method, lead to a summary of the Pareto-optimal selection systems. Both stages are discussed next. We illustrate the different steps of the procedure with an example application.

Step 1: Preparatory Analysis

Choice and articulation of selection goals. Generally, any selection decision should pay due attention to a careful analysis and articulation of the intended goals, thereby making a clear distinction between these goals and other concerns that are more properly defined as restrictions. In general terms, *goals* are objectives that one intends to optimize, whereas *restrictions* express minimal or maximal acceptable levels of an objective. To illustrate the distinction, the example application focuses on a planned selection system where both the goals of maximizing selection quality and work force diversity are valued. Although other concerns (e.g., the total selection cost) are also of importance, they will not be regarded as a goal that should be optimized but rather as a restriction that must be met by imposing a suitable set of constraints such as, for example, imposing a maximum acceptable bound on the total selection cost.

Once the selection goals are broadly defined in general terms (e.g., "quality" and "diversity" of the selected work force), the implementation of the analytic method to obtain Pareto-optimal selection systems is only modestly restrictive as to the choice of a metric for expressing levels of goal attainment. To illustrate the goal translation process, we return to the example introduced above. In the example, the selection quality goal is expressed as the expected mean job performance of the selected applicants, whereas the AIR gauges the different levels of goal achievement for the work force diversity objective.

Inventory of the available predictor battery and the expected applicant pool. In the second preparatory step, information is gathered on the available battery of screening procedures and the nature of the current or expected applicant pool. More specifically, this step results in data on the size and makeup of the applicant pool and values for the predictor validities, effect sizes, and intercorrelations. Note that all previous related work, whether its results are derived from equation (e.g., De Corte et al., 2006) or through simulation (e.g., Doverspike, Winter, Healy, & Barrett, 1996; Finch, Edwards, & Wallace, 2009), imposes identical requirements. In Appendix B, we address more concretely how users may obtain the data needed to apply our method. Issues related to uncertainty and sampling variability in the applicant and predictor data are studied in a next section of the article, whereas limitations related to the availability of the required data are further discussed in the final section of the article. It is important to emphasize, though, that the predictor data should reflect estimates of the validities, effect sizes, and intercorrelation values of the predictors at the applicant (and not the incumbent) population level, as selection is based on applicant groups rather than incumbent groups.

Table 1 details the predictor and applicant data used in the example application. The example focuses on designing Pareto-optimal selection systems for a planned selection where the available test battery consists of the five predictors (i.e., CA, SI, CO, BI, and IN) introduced in the opening section of the article. The table also includes the costs of administering the predictor per applicant as well as the values of the predictor effect sizes, intercorrelations, and validities (with respect to job performance), and it indicates that the candidate pool is a mixture of 88% majority applicants and 12% minority applicants. For the main part, the predictor data values displayed in Table 1 are drawn from Potosky, Bobko, and Roth (2005) and De Corte, Lievens, and Sackett (2008), whereas the values related to the IN predictor are based on results presented by Finch et al. (2009); McFarland and Ryan (2000); Ones and Viswesvaran (1998); Ones, Viswesvaran, and Schmidt (1993); Sackett and Wanek (1996); Schmidt and Hunter (1998); and Van Iddekinge, Raymark, Eidson, and Attenweiler (2004). We chose these data values because they represent current best estimates of the effect size, validity, and intercorrelation of the predictors at the applicant population level. We also drew the .88 and .12 majority and minority candidate proportions from Potosky et al. Note that the chosen predictor cost values are merely illustrative and primarily serve to show how cost data can be incorporated into the design process.

Table 1
Predictor and Applicant Data for the Example Application

Predictor	Cost (in dollars)	d[a]	1	2	3	4	5
1. CA	20	−0.72					
2. SI	150	−0.31	.31				
3. CO	20	−0.06	.03	.26			
4. BI	20	−0.57	.37	.17	.31		
5. IN	20	−0.04	.00	.00	.39	.25	
Criterion							
1. Job performance		−0.27	.51	.48	.22	.32	.41

Note. Applicant pool is a mixture of 88% majority and 12% minority candidates. CA = cognitive ability; SI = structured interview; CO = conscientiousness; BI = biodata; IN = integrity.
[a] The effect size *d* reports the standardize mean difference between the minority and the majority applicant groups.

Specification of the contextual constraints. The final preparatory step involves the specification of contextual features that govern the intended selection system. These contextual features typically impose restrictions on the set of feasible selection scenarios. For example, cost or logistical considerations may dictate that certain screening procedures, such as SI, can only be applied to a fraction of the initial applicants. Other considerations, reflecting practical constraints or organizational values, may result in adopting a non-compensatory selection rule and/or a partial sequencing of the predictors. This analysis of relevant contextual features can also result in total selection costs limits, a maximum value for the total number of administered predictors, an upper bound on the number of selection stages, upper and lower bounds for the retention rates in the intermediate selection stages in case of a multi-stage approach, and/or in restrictions on the way in which predictors are sequenced and weighted to the stage predictor composites. Obviously, the list of relevant contextual specifications may vary substantially from one application to the next, but the goal of this preparatory step is always the same: the exclusion of infeasible selection systems.

For the example application of our method, we consider three different choices for the set of relevant contextual constraints, each resulting in a particular selection situation. Table 2 summarizes the details of the three selection situations. The first situation, henceforth referred to as the *base line situation*, corresponds to the case where only a few contextual constraints are imposed, whereas the two other situations, labeled as *the cost constrained situation* and the *general constrained situation*, reflect increasingly constrained selection situations. As shown in Table 2, all three situations relate to the same battery of five available predictors and the same makeup of the expected applicant pool (cf. Table 1), and they all focus on the same planned selection with a .10 overall selection rate in which both the goals of maximizing selection quality and work force diversity are valued. However, the three situations differ in the specification of the set of relevant contextual constraints. Thus, Table 2 indicates that, in the *base line situation*, all five predictors will be administered, either in a single-stage, a two-stage, or a three-stage design. Also, the predictor weights may take any non-negative value when forming composites, and in the multi-stage designs, the retention rates at the end of the intermediate stage(s) are free to vary between 1 (i.e., no selection in the intermediate stage) and .10, the overall selection rate. Finally, to reflect common practice, usage of the SI predictor in two- and three-stage designs is limited to the final stage, and in case of a three-stage design, SI is the only permissible predictor in the third and final stage. Together, this limited set of contextual restrictions, as imposed in the base line situation, results in a total of 30 different acceptable selection scenarios. Table 3 gives an overview

Table 2
The Studied Selection Situations

The example application relates to three different selection situations: the base line situation, the cost constrained situation, and the general constrained situation.

Common contextual features

All three situations presume using all of the five predictors described in Table 1, either in a single-, a two-, or a three-stage design.

Also, all situations focus on the same planned selection with a .10 overall selection rate from an applicant group consisting of 12% minority and 88% majority group candidates.

The situations share the same two selection goals: maximizing the selection quality as expressed by the expected performance of the selected applicants and maximizing the diversity as gauged by the AIR.

Specific contextual characteristics

(a) Base line situation
- Non-negative predictor weights in forming stage specific predictor composites.
- Retention rates between .1 and 1 in the intermediate selection stages.
- Usage of the SI predictor is restricted to (a) single-stage designs, (b) the last stage of two-stage designs, or (c) be the only predictor in three-stage designs.

(b) Cost constrained situation
- The same specific contextual constraints as in the base line situation.
- The predictor cost per applicant may not exceed $100.

(c) General constrained situation
- The same specific contextual constraints as in the cost constrained situation.
- The CA predictor must be used in the first selection stage.
- The ratio of the predictor weights when forming composites may not exceed 5.
- In two-stage selection, the retention rate after the first stage must be between .25 and .50, whereas for three-stage selections, the retention rates must be between .40 and .80 (first stage) and between .20 and .50 (second stage).

Note. AIR = adverse impact ratio; SI = structured interview; CA = cognitive ability.

of these scenarios by detailing for each scenario the staging of the predictors.

Compared with the base line situation, the *cost constrained situation* adds a total predictor cost restriction to the situational constraints. More specifically, the situation requires that the total predictor cost per applicant may not exceed $100. Note that the cost constraint is added as an additional requirement and not as a third objective besides the goals of selection quality and work force diversity. As explained in the Further Issues and Extensions subsection, the latter translation of the cost concern as an additional objective is also possible. Also, we add the cost concern in the format of a total predictor cost per applicant because this translation permits the organization to control the total selection costs. Finally, note that adding the cost constraint does not eliminate using the SI predictor with an estimated cost of $150 but rather restricts its usage to only a fraction of the initial applicant group, as is the case in a multi-stage scenario.

For the final, *general constrained situation*, Table 2 shows that the contextual constraints express, besides the requirements of the cost constrained situation, that (a) the CA predictor must be used in the first stage; (b) the ratio between the largest predictor weight and the smallest predictor weight when forming predictor composites may not exceed 5; and (c) the retention rate after the first stage should lie between .25 and .50 in case of a two-stage selection, whereas the retention rates after Stages 1 and 2 must be between .40 and .80 and between .20 and .50, respectively, when adopting a three-stage selection design.

Although both the cost and the general constrained situation impose fairly realistic additional restrictions, it is emphasized that there is nothing sacrosanct about the values actually chosen for the total selection cost, the retention rates, and the maximum predictor weight ratio. The chosen values are merely illustrative and can be replaced by other values if appropriate. However, note that the limits on the predictor weights in the general constrained situation imply that all the predictors that are scheduled for use in a particular stage also effectively contribute to the predictor composite that is used in the stage. In addition, the retention rate limits adopted in this situation ensure that at least part of the applicants is screened out in the intermediate selection stages. Finally, we observe that the added contextual constraint in the cost constrained situation implies that the single-stage scenario using all predictors is no longer feasible, whereas the constraints of the general constrained situation further reduce the number of feasible scenarios to 11. The remaining 11 scenarios are labeled with a superscript "a" in Table 3.

Step 2: Computing a Representative Set of Pareto-Optimal Selection Systems

Given the selection situation resulting from the preparatory stage (e.g., the above specified base line situation), the present analytic method proceeds by solving for the associated Pareto-optimal selection systems and the corresponding Pareto-optimal trade-offs. Note that the derived systems are *globally Pareto-optimal* in the sense that they are Pareto-optimal across all the feasible selection scenarios and, therefore, need not all correspond to the same selection scenario. For the same reason, the corresponding set of Pareto-optimal goal trade-offs, which is usually referred to as the *Pareto front*, will often be a concatenation of subsets of Pareto-optimal trade-offs, with each subset corresponding to the implementation of one of the different feasible selection scenarios. For example, it may prove the case that a two-stage selection system proves Pareto-optimal for some portion of the AIR range, whereas a three-stage system proves Pareto-optimal in other portions of the range.

Obtaining the set of Pareto-optimal trade-offs is typically performed by means of multi-objective optimization methods. Following De Corte et al. (2007), we propose adopting the normal-boundary intersection method (Das & Dennis, 1998) to obtain a representative set of Pareto-optimal selection systems. Appendix A provides a detailed description of the method, including a summary of the assumptions required by its application. To implement the analytic method, a computer program—operating under the Windows operating system—was written. The program returns a summary description of the Pareto-optimal selection systems and a tabular display of the corresponding Pareto-optimal trade-off values for the selection goals. The operational details for implementing each system are provided as well. Appendix B provides a more detailed description of the practical aspects related to the imple-

Table 3
Selection Scenarios Studied in the Base Line Selection Situation

		Predictors used in the stages		
Scenario	No. of stages	Stage 1	Stage 2	Stage 3
1	1	CA, SI, CO, BI, IN		
2[a]	2	CA	SI, CO, BI, IN	
3	2	CO	CA, SI, BI, IN	
4	2	BI	CA, SI, CO, IN	
5	2	IN	CA, SI, CO, BI	
6[a]	2	CA, CO	SI, BI, IN	
7[a]	2	CA, BI	SI, CO, IN	
8[a]	2	CA, IN	SI, CO, BI	
9	2	CO, BI	CA, SI, IN	
10	2	CO, IN	CA, SI, BI	
11	2	BI, IN	CA, SI, CO	
12	2	CA, CO, BI	SI, IN	
13	2	CA, CO, IN	SI, BI	
14	2	CA, BI, IN	SI, CO	
15	2	CO, BI, IN	CA, IN	
16	2	CA, CO, BI, IN	SI	
17[a]	3	CA	CO, BI, IN	SI
18	3	CO	CA, BI, IN	SI
19	3	BI	CA, CO, IN	SI
20	3	IN	CA, CO, BI	SI
21[a]	3	CA, CO	BI, IN	SI
22[a]	3	CA, BI	CO, IN	SI
23[a]	3	CA, IN	CO, BI	SI
24	3	CO, BI	CA, IN	SI
25	3	CO, IN	CA, BI	SI
26	3	BI, IN	CA, CO	SI
27[a]	3	CA, CO, BI	IN	SI
28[a]	3	CA, CO, IN	BI	SI
29[a]	3	CA, BI, IN	CO	SI
30	3	CO, BI, IN	CA	SI

Note. CA = cognitive ability; SI = structured interview; CO = conscientiousness; BI = biodata; IN = integrity.
[a] Selection scenarios maintained in the general constrained situation.

mentation of the present decision aid and the accompanying computer software.

Results Example Application

To illustrate and discuss the results of the present analytic method, we applied the method to the three earlier introduced selection situations (cf. Table 2). We first report the global results obtained for each situation as captured by the graphical overview of the three corresponding global Pareto fronts and the description of the selection scenarios that contribute to these fronts. Next, we compare the results across the three situations and show how the application of the method also leads to a full description of the Pareto-optimal selection systems that populate the Pareto fronts.

When applied to the *base line selection situation*, our method results in the front of Pareto-optimal selection quality/diversity trade-offs as depicted by the dashed curve drawn in the upper panel of Figure 1. As suggested by the smoothness of this curve, the Pareto front corresponds to a single selection scenario even though the situation permits using all 30 scenarios. However, in one of these scenarios (i.e., the single-stage Scenario 1), all the predictor information can be used at once, in as optimal manner as possible, to strike the best possible balance in achieving the selection quality and diversity goals. In the remaining 29 scenarios, only part of the information can be used in initial stages of the selection system, implying that the selection quality/diversity potential of these scenarios cannot be better than the corresponding potential of the full information Scenario 1. So, in the base line selection situation, the set of Pareto-optimal selection quality/diversity trade-offs (i.e., the global Pareto front), associated with the entire set of 30 feasible selection scenarios, reduces to the set of Pareto-optimal trade-offs associated with Scenario 1. Inspection of the Pareto front shows that the Pareto-optimal trade-offs vary between a value of .932 for the AIR with a corresponding value of .689 for the selection quality objective (cf. trade-off point A) and an AIR value of .332 with a corresponding 1.267 value for selection quality (cf. Trade-Off B). Trade-Off A is the *diversity maximizing Pareto-optimal trade-off*, whereas Trade-Off B represents the *quality maximizing Pareto-optimal trade-off*. Compared with these two trade-offs, all other points on the Pareto front represent more balanced trade-offs between diversity and selection quality. Trade-Off C, with values of .643 and 1.080 for AIR and quality, respectively, offers an example of these more balanced Pareto-optimal trade-offs.

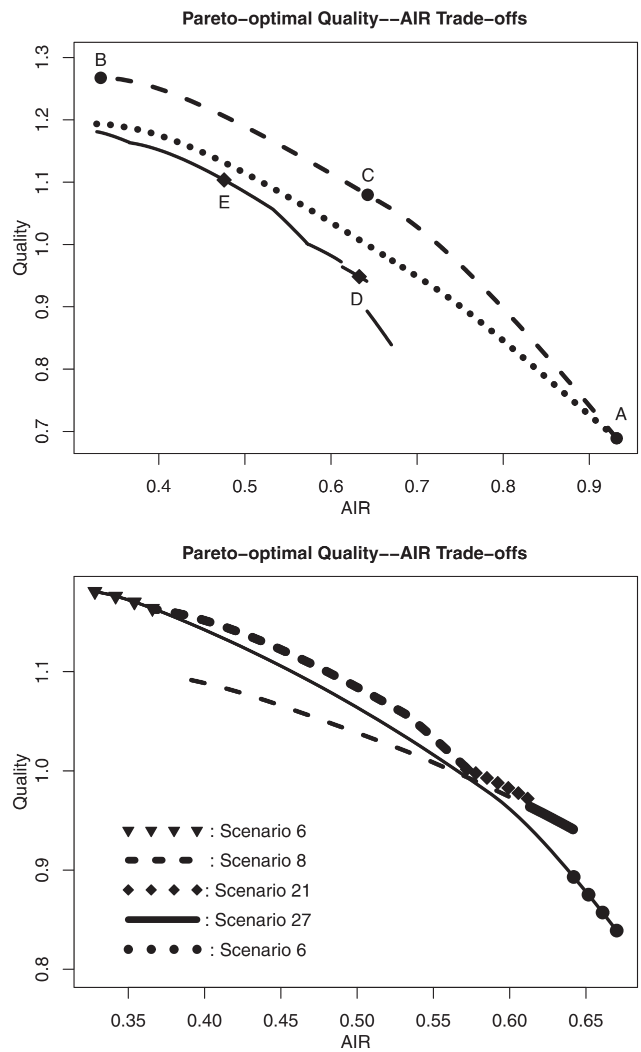

Figure 1. Upper panel: Pareto-optimal selection quality/work force diversity trade-off curves for the base line selection situation (dashed curve), the cost constrained situation (dotted curve), and the general constrained situation (solid line segments). Lower panel: scenarios contributing to the Pareto front of the general constrained selection situation. AIR = adverse impact ratio.

Next, we turn to the results of the *cost constrained situation*. Compared with the base line situation, this situation imposes the constraint that the total test costs per applicant may not exceed \$100. This constraint implies that Scenario 1, using all predictors in a single stage, is no longer feasible (as administering all predictors to the applicants results in a total test cost of \$230 per applicant). The other 29 scenarios remain possible, however, and the application of the method results in the global Pareto front, depicted as the dotted line in the upper panel of Figure 1. Again, the Pareto front looks smooth, even though several different scenarios contribute segments to the front. However, the different contributions are masked in the figure because the global Pareto front (i.e., the front as derived across all 29 feasible scenarios) is, for all practical purposes, identical to the Pareto front associated with the single Scenario 8. Over the entire range of the front, this scenario permits goal trade-off values that are either better than or almost identical to the best trade-off achievable with any other scenario.

Several interrelated features contribute to the superiority of Scenario 8 under the present cost constrained selection situation. First, in the first stage, the scenario uses two of the most valid predictors (i.e., CA and IN), one of which has the smallest (absolute) effect size value. Second, the situation permits applying any non-negative weighting of these predictors in forming the first-stage predictor composite. The resulting CA and IN composite can therefore lead to a broad range of favorable selection quality and AIR trade-off values, even under the rather severe first-stage selection rates that are required to cope with the overall test cost constraint. Third, again because it is permissible to use any non-negative weighting of the predictors in the composite formation, the selection quality and diversity trade-offs can be further improved in the second-stage selection of Scenario 8 by using composites of the remaining SI, CO and BI predictors that also show high validity (as contributed by the high-validity SI predictor) and low impact (due to the small effect sizes of both the SI and CO predictors). Finally, compared with other two-stage scenarios where three predictors are administered in the first stage, Scenario 8 permits more variability in the stage selection rates to meet the overall test cost constraint. With CA and IN, in the first stage, the overall cost constraint implies that the first-stage selection rate may vary between .10 and .32, whereas with any subset of three predictors from the predictors that are eligible for use in the first stage, the first-stage selection rate is limited to the interval between .10 and .24.

Finally, the solid line segments in the upper panel of Figure 1 present the Pareto front obtained for the *general constrained selection situation*. This time, the Pareto front is no longer smooth but shows several discontinuities indicating that the front corresponds to the concatenation of segments of the Pareto-optimal trade-off curves obtained under a number of different selection scenarios. This is further detailed in the bottom panel of Figure 1, where different line markings are used to identify the different segments of the Pareto front. Thus, the bottom panel of Figure 1 shows, from upper left to lower right, that the Pareto front is built from segments of the Pareto-optimal trade-off curves associated with Scenarios 6, 8, 21, 27, and 6.

Observe that discontinuities in the global Pareto front are not an anomaly. Instead, they translate the effect of the additional contextual constraints (cf. Table 2, Subsection c) that define the general constrained selection situation. Although Scenario 8 still contributes the largest, central segment to the global Pareto front, the additional constraints imply that other scenarios (i.e., Scenarios 6, 21, and 27) have a better potential for generating quality/diversity trade-offs that are characterized by either a low or a high AIR value. Thus, Scenario 6 contributes both the upper left and the lower right extreme part of the global Pareto front, whereas Scenarios 21 and 27 add segments with quality/diversity trade-offs that are characterized by a rather high value for the AIR. The potential of Scenario 6 in generating the upper left Pareto front segment is driven by using maximum validity composites (within the limits imposed by the predictor weighing constraint) in the two selection stages, whereas the contribution of Scenarios 6, 21, and 27 to the lower right segments of the global Pareto front relates to two interacting features. First, these scenarios all use CA (a high-validity, high-impact predictor) and CO (a low-validity, low-impact predictor) in the first stage of the selection process, typically assigning a substantially higher weight to the CO predictor and resulting in a moderate validity but a high AIR trade-off after the first stage. Next, the validity is substantially increased, without sacrificing the AIR, by applying a second-stage composite (with a high level of selectivity) that is either dominated by the high validity/low impact predictor IN (Scenarios 6 and 21) or identical to the IN predictor (Scenario 27).

As expected, the comparison of the Pareto fronts obtained for the base line and the cost constrained situation reveals that the total test cost constraint limits the quality of the attainable Pareto-optimal trade-offs. Over a broad range of AIR values (i.e., for AIR values between .33 and .80), the quality trade-off values of the base line situation are between .05 and .10 units higher than their counterpart values in the cost constrained situation. Although these differences may seem small, it is important to observe that they express differences in expected level of job performance that, when translated in the dollar metric, may be quite substantial compared with the gain achieved by imposing the total test cost constraint.

Compared with the Pareto fronts achievable in the base line and the cost constrained situations, the additional contextual restrictions imposed in the general constrained situation further affect the quality of the attainable Pareto-optimal diversity/validity trade-offs. In particular, these added constraints exclude Pareto optimal trade-offs that show higher AIR values (i.e., for AIR values greater than .67). This effect is largely due to the constraint on the permissible ratio of the predictor weights when forming predictor composites. Because of this constraint and the requirement to use the CA predictor in the first stage, first- and second-stage composites—which are almost exclusively dominated by a low impact predictor such as the CO and the IN predictor—are no longer admissible. As a consequence, the intermediate stage composites all show moderate to high effect sizes, which, in combination with the restriction on the acceptable retention rates in these stages, inevitably leads to the exclusion of quality/diversity trade-offs with higher AIR values.

Besides a summary of the global quality/diversity trade-off potential of the studied selection situations and the identification of the selection scenarios that contribute to the Pareto optimal front, the application of the method also results in detailed information for the design of the selection systems that correspond to the Pareto-optimal trade-offs. This design information effectively specifies how each of the optimal selection systems can be implemented. As an example, consider the Pareto-optimal trade-off curve depicted in the upper panel of Figure 1 for the *general constrained selection situation* and Points D and E on this curve, in particular. Trade-Off D, with quality and AIR values of 0.948 and 0.633, respectively, results from Scenario 27, using the CA, CO, and the BI predictors with weights of .08, .32, and .06 in the first stage, retaining 50% of the applicants; the IN predictor in the second stage, retaining 20% of the initial applicants; and the SI predictor in the final stage. Alternatively, Trade-Off E (with quality and AIR values of 1.103 and 0.476, respectively) is obtained when implementing Scenario 8 with weights of .18 and .52 for the first-stage composite of the CA and IN predictors, retaining 25% of the applicants, and weights of .46, .09, and .09 for the second-stage composite of SI, CO, and BI. Note that the set of predictor weights for each Pareto-optimal trade-off is part of the output produced by the program; we do not report the weights for each point on the Pareto frontier, as they are specific to this example and not of general interest.

Discussion

Despite the difference in the general level of the Pareto fronts obtained under the three analyzed selection situations, the results displayed in Figure 1 first and foremost illustrate the substantial diversity of Pareto-optimal selection trade-offs that can be achieved under each situation. In all three situations, it is perfectly possible to implement selection systems that result in a high work force diversity, other systems that promote high levels of adverse impact, and still others that balance the diversity and quality concerns. In the three example situations, the selection practitioner has a very substantial range of options to choose from, and one may expect this to be also the case in many other settings. To the extent that the battery of available predictors contains a mix of high-validity and low-impact predictors and the contextual constraints permit using predictor subsets that reflect this variety, the gamut of attainable Pareto-optimal selection quality/work force diversity trade-offs will typically be quite substantial.

A second important issue, already noted by De Corte et al. (2007) when discussing the design of optimal predictor weighting systems, is that the present method does not lead to a single optimal solution unless the selection goals are prioritized in the sense that achievement on a lower ranked goal becomes important only after maximization of the higher ranked goals. In that case, only the Pareto-optimal solution that corresponds to the goal hierarchy is relevant. Thus, if one seeks to first maximize selection quality (diversity) in the example application, then only the quality (diversity) maximizing Pareto-optimal solution is acceptable. Alternatively, in the absence of goal prioritization, all Pareto-optimal solutions are of potential interest, and it then falls to the selection practitioner to evaluate the relative merit of the corresponding Pareto-optimal trade-offs and to ultimately decide in favor of one of the trade-offs on the basis of additional considerations that reflect a value statement on the particular kind of balance between selection quality and work force diversity one is aiming at (cf. De Corte et al., 2008). As noted by these authors, this feature should not be regarded as a weakness of the method. Instead, it reflects the mere fact that "the resolution of competing goals typically requires a decision as to the relative importance of these objectives" (De Corte et al., 2007, p. 1387).

Although the present method does not single out one particular optimal selection system, it nevertheless clearly differentiates between systems that deserve further consideration (i.e., the Pareto-optimal ones) and those that do not. Figure 2 illustrates this point by providing a complete overview of all the quality/diversity trade-offs that are achievable for the example application under the *general constrained selection situation*. Each point within the closed contour on the figure indicates an achievable trade-off and corresponds to a particular choice for the design of the selection system. Yet, all the selection systems with an associated trade-off that does not lie on the Pareto front indicated by the thick line upper parts of the contour are suboptimal and should therefore not be considered for further implementation.

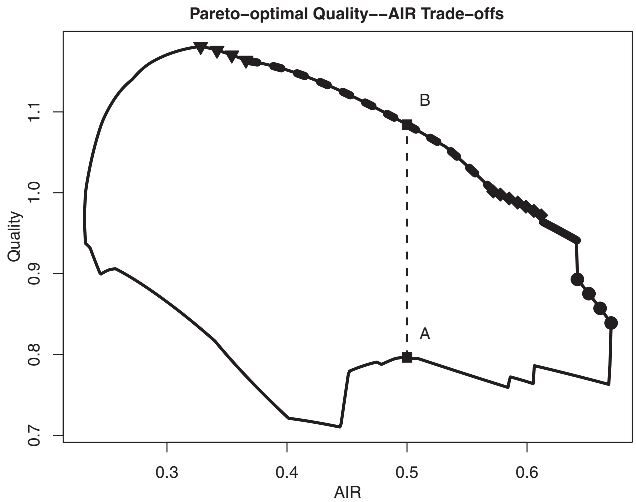

Figure 2. Gamut of all achievable selection quality/work force diversity trade-offs under the general constrained selection situation. The achievable trade-offs are within the enclosure formed by the thin solid line segments and the bold upper right line segments. AIR = adverse impact ratio.

Figure 2 also demonstrates the huge importance of designing selection systems properly. Following the same approach as De Corte et al. (2006, 2008), a vertical line is drawn within the achievable trade-off contour to emphasize the detrimental effect of improper selection system design. The vertical line shows that for one AIR value selected for illustrative purpose (i.e., .50), in the worst case one may end up with a system that results in an expected average job performance of only .79 (cf. Point A in Figure 2) compared with the corresponding Pareto-optimal selection design that for the same AIR offers an expected job performance of 1.08 (cf. Point B)—an almost 40% increase.

Thus far, the discussion of the results focuses on the global Pareto-optimal selection systems (i.e., systems that are Pareto-optimal across all selection scenarios that are feasible for a given selection situation) and the corresponding global Pareto front of quality/diversity trade-offs. This is the obvious choice when one intends designing selection systems that are expected to perform optimally for the given selection situation because this implies seeking the best possible systems over all selection scenarios that are admissible within the prescribed selection situation. However, note that the application of the method also results in the identification of the selection systems and the quality/diversity trade-offs that are Pareto-optimal with respect to each specific feasible selection scenario (cf. Appendix A). As an example, consider the thin solid and the thin dashed line added to the lower panel of Figure 1. These lines depict the Pareto front corresponding to Scenarios 6 and 27, respectively, and they are added to the figure to show the relation between the Pareto-optimal front as obtained over the entire set of feasible scenarios (i.e., the global Pareto front) and the Pareto fronts associated with particular feasible scenarios.

Further Issues and Extensions

To improve the focus and flow in the presentation of the method for designing Pareto-optimal selection systems, a number of issues—related to the applicability of the method and the choice and translation of the selection goals—remained out of scope. For the same reasons, we also refrained from discussing related extensions that generalize the applicability of the method. We elaborate on these issues and the corresponding extensions in the next paragraphs.

First, we note that, although immediately applicable to fixed applicant pool selection systems (i.e., systems where the pool of candidates is fixed prior to any screening of candidates), the method is also applicable to continuous flow selection systems (i.e., settings in which decisions need to be made about each individual applicant, rather than waiting for a formal candidate pool to be specified) as long as the size of the applicant pool can be estimated and the overall selection ratio for the selection system can be specified. In other words, if it can be specified, say, that 5% of an estimated 40,000 candidates per year is to be selected, the method applies to both continuous flow and fixed pool systems.

Second, the example application focused on selection decisions with a single job performance dimension, the presence of only one minority group, and the pursuit of only two selection goals. These limitations were for illustrative purposes only; the approach and the method for implementing it permit more than one job performance dimension, several minority applicant groups, and more than two selection goals.

Third, regarding the operationalization of the selection goals, we note that any metric that preserves the ordinal properties of goal achievement is acceptable because metrics that share this property, but are otherwise different, lead to the same set of Pareto-optimal selection systems. Thus, using the AIR metric to translate the work force diversity goal, or choosing the Fisher Exact or the Z_{IR} test probabilities as done in Finch et al.'s (2009) study, would not affect the resulting set of Pareto-optimal selection systems because both these test probabilities and the AIR are for any given total number of applicants monotonically increasing functions of the minority group selection rate. The consequence is that any selection system that is Pareto-optimal in terms of expected average job performance and AIR will also be Pareto-optimal in terms of expected average job performance and, for example, the Fisher Exact test probability. Therefore, the resulting Pareto-optimal selection systems are invariant for any other choice of metric that is monotonically related to the present operationalizations.

Notwithstanding the above remarks, one may still question whether aiming to reduce adverse impact and desiring to increase diversity are indeed interchangeable on the grounds that the pursuit of diversity is a more embracing objective that also relates to specific recruitment efforts and goes beyond reducing adverse impact, associated with selection tests, against a specific minority group. Although these are valid reflections, we believe that the AIR provides an adequate metric for the diversity goal, at least in the present context where the focus is on designing optimal selection systems *given the recruited applicant pool and the available predictors*.

Generalizability of Pareto-Optimal Selection Systems

The Pareto-optimal selection systems and the corresponding Pareto-optimal trade-offs obtained by the present decision aid depend on the values used for the predictor and the applicant data. However, in applications where the decision aid is used to design a future planned selection, only estimates instead of the actual data values will typically be available. It is therefore important to assess whether Pareto-optimal systems, as obtained from estimated data input values, continue to perform well under future realizations of the predictor and applicant data. To address this issue we performed two simulation studies. The first study adopts a *population to sample approach* by investigating the sensitivity of Pareto-optimal selection systems to sampling-based departures from the population data input values. Here, Pareto-optimal systems derived from population input data are evaluated in terms of how well they work when applied to samples from that population. The second study adopts a *sample to population approach* (cf. the population cross-validation approach described by Schmitt & Ployhart, 1999) and compares the quality of Pareto-optimal selection systems, as derived from sample data, with the quality of the Pareto-optimal systems, as based on population input data values. In essence, this is the converse of the first investigation. Whereas that investigation examined whether solutions that were Pareto-optimal in the population would also perform well in samples from that population, the second investigation identifies solutions that are Pareto-optimal in samples and examines whether those solutions would perform well in the population from which the samples were drawn.

Population to Sample Perspective: Sensitivity Study

Procedure. To study whether Pareto-optimal selection systems, as derived from meta-analytic population input data estimates, continue to perform well when applied to samples drawn from that population, we adopted an improved version of the sensitivity analysis originally proposed by De Corte et al. (2007). The analysis starts with the computation of a number of Pareto-optimal selection systems as obtained from the population input data values. For each Pareto-optimal selection system, we then generate a corresponding set of inferior, non-optimal selection systems. We term these as systems that are "dominated" by the optimal system because they are outperformed on the outcomes of interest by the optimal system.

Next, simulation methods are used to construct 1,000 samples based on the initial population input data values. However, in contrast to De Corte et al. (2007), the present procedure to construct the sample data input values employs more sophisticated simulation schemes that tie the level of sampling variability to the total number, n, of candidates in the applicant pool. In particular, the procedure described by Hong (1999) is used to generate the samples of the predictor correlation and validity data values, whereas the effect sizes are obtained by sampling from normal distributions with mean equal to the initial effect size estimate d and variance equal to $n/(n_1 + n_2) + d^2/2n$ (cf. Hedges & Olkin, 1985), with n_1 and n_2 representing the number of majority and minority candidates in the applicant sample.

Finally, each generated sample is used as input to compute the trade-off values associated with the Pareto-optimal selection systems and each of the dominated selection systems. Following the proposal by De Corte et al. (2007), the sensitivity of a Pareto-optimal selection system for sampling variability in the data input values can then be gauged by two sensitivity indices. The first is the percentage, over all 1,000 samples and all dominated selection systems, in which the "right" answer is reached, namely that the system that is optimal in the population is found superior to systems that are not optimal in the population. The second is the converse: the percentage of comparisons in which the "wrong" answer is reached, namely, that a system that is non-optimal in the population is found superior to the optimal system. These two do not sum to 100%, as there are comparisons in which neither trade-off dominates the other. Also note that the two indices have values equal to 100% and 0% if a Pareto-optimal selection system would be completely insensitive to sampling variability in the data input values, whereas under complete sensitivity, both indices are equal to 25% (i.e., in settings where one system exceeding its comparison on the two outcomes of AIR and mean performance is a random event, the chances of one exceeding the other on both outcomes are 25%).

As discussed in Appendix A, the present decision tool allows for several choices as to the weighing of the predictors when forming predictor composites. For either unit, regression-based or optimal weighing of the predictors is possible. In light of the long standing debate on the usefulness of unit weights compared with differential weights when forming predictor composites (cf. Bobko, Roth, & Buster, 2007), we applied the above described procedure to two sets of Pareto-optimal selection systems. Both sets are computed for the same general constrained selection situation, using the same population-based data input values, but the Pareto-optimal selection systems of the second set also comply with the additional requirement that only unit weighted predictor composites are acceptable. This enabled us to compare the potential of optimally weighted (Set 1) and unit weighted (Set 2) Pareto-optimal selection systems in conserving more favorable quality/diversity trade-offs than the trade-offs associated with their corresponding dominated systems under realistic conditions of sampling variability in the applicant and predictor data.

Results. Columns 2–5 of Table 4 summarize the results of the sensitivity analyses as obtained for total applicant sizes of 100,

Table 4
Results Sensitivity and Cross-Validation Simulation Studies

	% of comparisons in which Pareto-optimal system dominates non-optimal system		% of comparisons in which non-optimal system dominates Pareto-optimal system		Relative quality of sample-based Pareto-optimal system relative to population-based system	
Total applicant size	Optimal weights	Unit weights	Optimal weights	Unit weights	Optimal weights	Unit weights
100	60.4	59.9	3.1	3.3	.76	.83
250	72.5	73.4	0.8	0.9	.84	.91
500	80.1	81.3	0.2	0.2	.88	.94

250, and 500, respectively. Columns 2 and 3 show the percentage with which the Pareto-optimal selection systems, as based on population data input values, continue to outperform their dominated selection systems under sampling variability of the input values, whereas columns 4 and 5 show the percentage of comparisons in which these Pareto-optimal systems become inferior to their dominated systems.

The obtained results uncover two basic findings. First, Pareto-optimal systems as computed from meta-analytic (population) applicant and predictor input data continue to perform well even when the sampling variability in the data input corresponds to the variability that can be expected for rather small total applicant pools (i.e., total applicant sample equal to 100). Even for such small applicant pool situations, one may expect that the population-based Pareto-optimal selection systems outperform their associated dominated systems in roughly 60% of the cases, which is considerably larger than the 25% expected by chance. Also, the corresponding percentages that the Pareto-optimal systems will be dominated are all very small. Second, Pareto-optimal systems that allow for optimal predictor weighing when forming predictor composites and Pareto-optimal systems with unit weighted composites hardly differ in terms of sensitivity for sampling variability in the data input values.

Sample to Population Perspective: Population Cross-Validity Study

Procedure. The second simulation study assesses the performance of Pareto-optimal systems, computed from sample-based data input values, relative to the performance of Pareto-optimal systems derived from the population input data. To achieve this purpose, we used a four-step procedure that mimics the approach adopted in the predictive regression context (cf. Schmitt & Ployhart, 1999). In the first step, we generated 1,000 samples for the data input values from the Table 1 population data values using the approach described in the previous Procedure subsection. Next, we implemented our method to determine a representative set of Pareto-optimal systems associated with each of the 1,000 data input samples. In the third step, we applied these systems to the population data input values. This resulted for each data input sample and each system in a quality/diversity trade-off value that expresses the trade-off expected for the system when applied to the population input data. Finally, we compared these trade-off values with the Pareto-optimal trade-offs achievable under the population data input conditions.

We repeated the above procedure for three different levels of sampling variability in the data input values, reflecting selections with 100, 250, and 500 applicants, respectively. Also, similar to the previous study, the present simulation again relates to the general constrained selection situation, and all analyses are performed both for the optimal predictor weighting and the unit weighting conditions.

To index the performance of the sample-based Pareto-optimal systems, at each value of the AIR, we computed the ratio of the level of the performance objective obtained in the sample with that obtained for the population input data, and we averaged this ratio across levels of AIR. Thus, for each sample, this index can be interpreted as how close the sample-based system comes to the quality of the Pareto-optimal solution that one would obtain if one had access to population data. The index takes a value of 1.0 if the sample was to match the population in the quality of the solution. Averaging this index across samples for each examined sample size gives a summary value as to the degree to which sample-based solutions approximate population solutions.

Results. The sixth and seventh columns of Table 4 report the "population cross validity" (i.e., the relative quality) values obtained for the optimal and the unit predictor weighing conditions under the three studied levels of sampling variability in the data input values. To interpret these values, it is noted that the expected value of the present index—over all feasible selection systems that result in the same diversity but a different quality value—equals .50. The presently obtained values, ranging between .76 and .94, therefore indicate that sample-based Pareto-optimal selection systems show substantially higher population cross validity than arbitrary designed feasible selection systems, even when the systems derive from small sample-based data input values. However, unlike the previous simulation, the present study suggests that Pareto-optimal selection systems using unit predictor weighting perform somewhat better than the corresponding systems using optimal predictor weighing.

Discussion

In general, the present results confirm and generalize the earlier finding of De Corte et al. (2007). In both studies, the results of the sensitivity analysis indicate that methods aiming for Pareto-optimal predictor weighing systems (cf. De Corte et al., 2007) and more general Pareto-optimal selection systems (cf. the present study) also perform well when the actual predictor and applicant data reflect sampling-based uncertainty. Although the present results relate to only one problem situation involving a set of five predictors, we repeated the sensitivity analysis for a number of other situations and obtained results that are very similar to the present findings, thereby confirming the superiority of Pareto-optimal selection systems over other designs when several selection goals are judged important.

The results of the second simulation study further indicate that Pareto-optimal selection systems also show substantial levels of population cross validity, but in contrast to the first simulation, this study also indicates that Pareto-optimal systems using unit predictor weighting cross validate better than corresponding systems using optimal predictor weighting. This is not an unexpected finding given similar observations in the simple prediction context, however.

Addressing Research Issues on Managing the Quality/Diversity Trade-Off

As the simultaneous pursuit of high quality and low impact selections remains one of the most vexing challenges in personnel selection, it is no wonder that several studies have focused on specific recommendations for better managing the quality/diversity trade-off. Although these studies met with some success for designing single-stage and limited two-stage selection systems, and despite the recent efforts of Finch et al. (2009) to proceed with more complex design, we concur with Sackett and Roth (1996) and De Corte et al. (2006) in the opinion that "there are no simple rules that can be offered about which approach to hurdle-based selection is preferred" (Sackett & Roth, 1996, p. 569). In fact, given the present method for designing optimal selection systems, we believe that the search for rules of thumb that go beyond sheer logic (e.g., if possible, apply equally valid but lower impact predictors; cf. De Corte et al., 2006) is no longer a primary concern because the method reported here routinely resolves the major operational selection design issues. In addition, it is shown in the next subsections that the method offers a tool of choice for addressing a number of more general selection research questions. In particular, we study four issues: the impact of using different predictor weighing systems, the comparison of compensatory and noncompensatory systems, the comparison of single-stage versus multi-stage systems, and the consequences of the practice of using composites of only predictors administered within one stage of a selection system rather than also incorporating predictor information from earlier stages.

Impact of Using Different Predictor Weighing Systems

Several previous studies addressed the impact of using different predictor weighing systems. However, all these studies were limited to studying this effect for a given set of predictors as applied within a particular selection scenario. These studies could therefore not assess the importance of the predictor weighting effect within the broader context of simultaneous predictor subset, selection rule, selection staging, and predictor sequencing decisions. Yet, the literature suggests that these latter decisions may be substantially more important in reaching selection outcomes that show both high quality and low impact (cf. Ployhart & Holtz, 2008).

To address the issue, the present method was used to compute the Pareto front for the example application employing either unit or regression weights, instead of the previous optimized weights, when combining predictors to predictor composites. All computations are performed with respect to the *cost constrained selection situation* examined in prior sections of the article. Figure 3 summarizes the results of the computations. The upper panel of the figure compares the Pareto front obtainable using optimal predictor weighting (dotted line) and regression weighting (solid line), respectively, whereas the bottom panel provides the same comparison when using unit predictor weighting. Note that, as before, all depicted fronts represent global Pareto optimal fronts as obtained over the 29 scenarios that are feasible in the cost constrained situation (cf. Table 3).

Both displays emphasize the superiority of the optimal predictor weighting approach for obtaining favorable selection quality/adverse impact trade-offs. So, predictor weighting remains an important issue even when accounting for variability in the selection staging, the predictor sequencing, and the stage retention decisions. However, the displays of Figure 3 also indicate that the incremental contribution of choosing optimal weights varies over the range of attainable values for the AIR. For AIR values in the range of .68–.92, the three weighting strategies result in almost the same Pareto front, whereas more substantial discrepancies, in favor of the optimal weighting approach, occur for AIR values between .35 and .68.

The finding that regression weighting can lead to lower expected performance than Pareto-optimal weighting may seem counterintuitive, given the general expectation of the superiority of regression weights. However, the key point is that regression weights will be superior in the absence of any other constraints. If one simply seeks to maximize performance among those selected, regression weighting will indeed be optimal. However, in the face of additional constraints (e.g., with a constraint on cost per applicant), other weights can dominate regression weights. This superiority often derives from the fact that, for a given AIR value, selection scenarios, which otherwise (i.e., with optimal predictor weighting) lead to Pareto-optimal solutions, can no longer result in this AIR value when using regression weights. Thus, whereas the Pareto-optimal trade-offs in the optimal weighting condition typically result from applying Scenario 8, this scenario cannot be used in the regression weighting condition to obtain Pareto-optimal trade-offs with an AIR value greater than .32 because the regression composite of the two predictors used in the first stage of the scenario (i.e., the predictors CA and IN; cf. Table 3) results in very

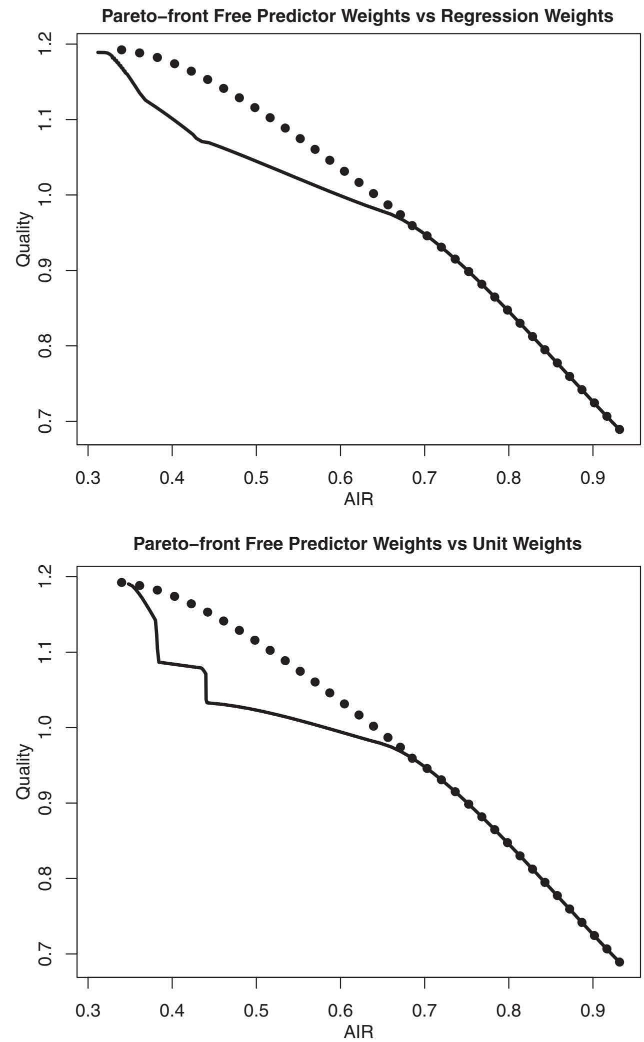

Figure 3. Impact of using different predictor weighing systems when forming predictor composites in the cost constrained selection situation. Upper panel: optimal predictor weights (dotted line) versus regression weights (solid line). Lower panel: optimal weights (dotted line) versus unit weights (solid line). AIR = adverse impact ratio.

low AIR values that cannot be compensated for in the second stage of the selection.

On a more general level, the irregular pattern of discrepancies between the Pareto-optimal trade-offs obtained under optimal predictor weighting and those obtained under regression or unit weighting further confirms that the pursuit for rules of thumb about the effect of different predictor weighing scenarios does not look very promising. Instead, it seems more fruitful to apply the present procedure to summarize the potential of the different weighting options, thereby creating the opportunity for a much better informed decision on the preferred weighting system than is otherwise possible.

Comparing Compensatory and Non-Compensatory Selection Rules

Virtually all previous research on the quality/diversity merits of alternative selection approaches focused on compensatory selection designs. However, in actual practice, it regularly happens that the score on one or more predictors cannot be compensated for by the scores on other predictors. Using cutoff scores, based on job analysis or reflecting legal or institutional conventions, represents a fairly generic example of this practice. Because of the paucity of relevant research, knowledge of the potential effects of adopting a non-compensatory selection rule on the achievable quality/diversity trade-offs is limited to a few logic-based expectations. Thus, it is expected that the effects on diversity will be larger when the non-compensatory predictor shows higher effect size values and is used with higher cutoff values. In addition, it is expected that cutoff values on high-validity predictors will result in higher selection quality. However, logic derivation by and large fails when addressing the simultaneous effects on quality and diversity of adopting a non-compensatory selection rule in the fairly typical case of a planned selection system using several predictors. In these situations, we again recommend using our method to explore the effects on the valued selection goals of implementing the planned non-compensatory rule and to compare these effects with those expected for the corresponding compensatory approach.

To illustrate our proposal, we applied our method to the *general constrained selection situation* assuming a non-compensatory rule for the CA predictor scores. In particular, we determined the Pareto front for this situation under the additional contextual constraint that only candidates who have a standardized (with respect to the majority candidate population) CA score at least equal to –1 can be selected (single-stage scenario) or retained for further scrutiny (multi-stage scenarios). The upper panel of Figure 4 depicts the resulting quality/diversity Pareto front (cf. the solid line segments) together with the Pareto front obtained under the corresponding compensatory situation (cf. the dotted line segments). The results indicate a very substantial decrease in the quality/diversity trade off potential when adopting the non-compensatory selection rule. The range of the attainable Pareto optimal trade-offs is severely restricted compared with the compensatory situation, and all attainable Pareto optimal trade-offs are dominated by the corresponding trade-offs under compensatory selection.

Using a CA cutoff value of –1 in the above example effectively means cutting slightly below the average CA value of the minority applicant population (equal to −.72; cf. Table 1) and therefore represents a rather severe hurdle for these applicants. However, the effects of the non-compensatory rule for using the CA predictor remain substantial even for less severe values of the cutoff. Relaxing the cutoff score from –1 to, for example, –2, leads to the Pareto front depicted by the solid line segments in the lower panel of Figure 4. Although the front offers a wider range, in terms of the AIR metric, than the corresponding front under the –1 CA cutoff situation, it still has many Pareto-optimal trade-offs that are dominated by trade-offs obtainable in the compensatory approach. Also, the Pareto front under the non-compensatory selection rule using a −2 CA cutoff value is shifted to the right compared with the corresponding front under the –1 CA cutoff situation. This shift, expressing that similar levels of selection quality are achievable for higher AIR values, relates to the substantial effect size of the non-compensatory CA predictor. With a lower cutoff value of

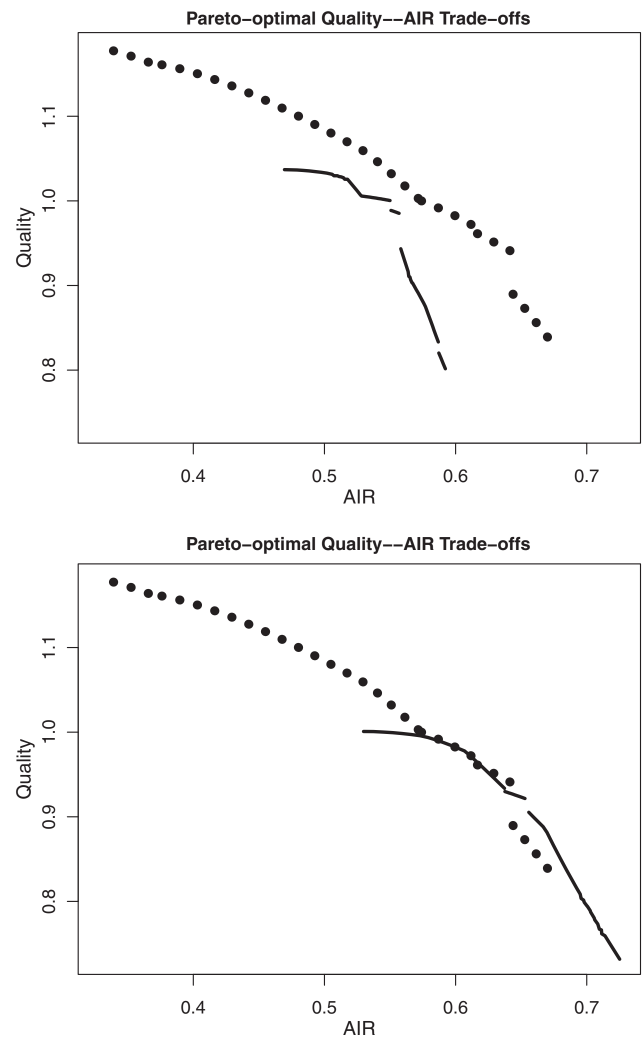

Figure 4. Comparing compensatory versus non-compensatory selection rules for the general constrained selection situation: selection quality/diversity Pareto front under compensatory selection (dotted curve), non-compensatory selection for CA with cutoff equal to −1 (solid lines, upper panel), and non-compensatory selection for CA with cutoff equal to –2 (solid lines, lower panel). AIR = adverse impact ratio.

–2, the CA predictor is used less selectively, resulting in less adverse impact. Also, according to the same reasoning, it follows that the direction and the magnitude of the shift will in general depend on the effect size of the non-compensatory predictor.

In summary, the above application shows that deciding in favor of a non-compensatory selection system may have a serious effect on the attainable Pareto front. As argued above, the magnitude of the effect relates to the effect size and the particular cutoff value chosen for the non-compensatory predictor(s). However, the extent of the effect will also vary according to, among others, the characteristics of the other available predictors, the overall selection rate, the eventual staging of the non-compensatory predictor(s), and the contextual constraints governing the selection situation. Because of this virtually unlimited variability in selection conditions, we highly recommend applying our method, given the relevant contextual specification of the planned selection, to gauge the effect on selection quality and diversity of adopting a non-compensatory selection rule.

Comparing Single-Stage Versus Multi-Stage Selection Designs

Previous research points to the potential that multi-stage selection designs may have over single-stage models for reducing AI. In fact, Finch et al. (2009, p. 330) observed "that multi stage strategies will generally produce less AI than similarly comprised single stage strategies" provided that the lower impact predictors receive due weight in the selection process. These authors also found that multi-stage strategies are more effective than single-stage strategies in balancing the goals of selection quality and work force diversity. Obviously, these conclusions cannot hold in general. In the first application of our method, investigating the merits of alternative selection scenarios under the *base line selection situation*, it was argued and subsequently verified that the Pareto front of the single-stage design dominates the Pareto front associated with the multi-stage designs. Disregarding total predictor costs and eventual logistical considerations, single-stage strategies are normally expected to result in better Pareto-optimal quality/diversity trade-offs than multi-stage scenarios that use the same predictor set because single-stage scenarios can use all available predictor information at once in striking Pareto-optimal balances, whereas multi-stage scenario can use only parts of this information in succession.

Whether multi-stage designs permit better Pareto-optimal quality/diversity trade-offs than single-stage approaches ultimately depends on the contextual constraints that govern the selection situation at hand. Thus, it may happen that these contextual constraints imply that the single-stage design is infeasible as is the case in, for example, the earlier introduced *cost constrained selection situation*. In other situations, both single- and multi-stage designs may be feasible where the multi-stage designs outperform the single-stage approaches. As an example, consider the *general constrained situation*, with the modified constraint that single-stage designs using only four predictors are acceptable. In that case, the single-stage scenario—using all but the SI predictor—becomes feasible, and our method can be applied to each of the sets of single-stage, two-stage, and three-stage scenarios separately to obtain the Pareto-optimal fronts that correspond to the three different levels of selection staging. Figure 5 reports the results of the application: The circle-shape curve corresponds to the Pareto front for the single-stage design, the triangle-shape line segment represents the Pareto front for the two-stage designs, and the square-shape curve shows the Pareto front of the three-stage designs. This time, the display corroborates the earlier claim that a multi-stage design is usually more effective than the single-stage approach in achieving a better balance between selection quality and adverse impact. Except for extreme values of the AIR (i.e., values lower than .32 and above .67), the Pareto fronts of the multi-stage selection strategies dominate the Pareto front of the single-stage approach, with the three-stage strategy offering the largest variety of Pareto-optimal trade-offs.

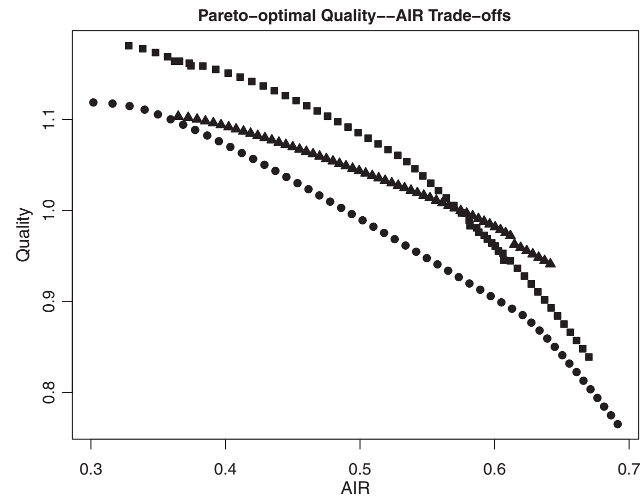

Figure 5. Comparing single-stage versus multi-stage designs for the modified (see text) general constrained selection situation: selection quality/diversity Pareto front of single-stage (bullet curve), two-stage (triangle curve), and three-stage (diamond curve) designs. AIR = adverse impact ratio.

As planned selection systems may vary considerably in terms of the available predictors and their characteristics as well as in terms of the prevailing contextual constraints, any general advice on preferring single- or multi-stage strategies is a dead alley, however. Instead, we again recommend using our method as was illustrated above to explore the potential of the different staging approaches, given the actual available predictors and the applicable contextual constraints, and to decide the issue on the basis of the thus obtained results.

Using Only Stage-Administered Predictors in the Composite Formation

It is common practice to base intermediate and final selection stage decisions on composite scores from the stage-specific administered predictors, disregarding previously obtained predictor information. A possible reason for this practice may be the concern that otherwise the composites will be highly correlated, resulting in intermediate and final-stage decisions that are all based on very similar predictor composites. However, this reason as well as others—such as, for example, the concern that AI will become all the worse by using information from high impact predictors in several consecutive stages—may be ill-founded when the predictor weighting in the stages is not fixed a priori. In such cases, valuable predictor information obtained in the earlier stages is neglected in the later stages, eventually resulting in selection systems that show less potential for balancing quality/diversity concerns than would otherwise be possible.

To study the issue, we compare the earlier derived Pareto front for the *general constrained selection situation*, obtained using only stage-specific administered predictor information, with the corresponding front associated with using predictor composites from both the stage-specific and the earlier administered predictors. The upper panel of Figure 6 summarizes the two Pareto fronts: The solid curve corresponds to the stage-specific predictor usage condition, whereas the dotted curve represents the Pareto front under the all available predictor usage situation. In contrast to initial intuition, the figure shows that using all available predictor data, either obtained in the stage itself or a previous one, does not always lead to Pareto-optimal quality/diversity trade-offs that outperform the corresponding trade-offs when only the stage-specific predictor information is used in forming the stage predictor composites. However, the results are consistent with a more thorough analysis of the situation. This situation (i.e., the *general constrained situation*) stipulates using the high-validity and high-impact CA predictor in the first stage. The predictor is therefore used repeatedly when all present and previous predictor information contributes to the stage composites, implying an increased validity of the composites and a higher quality of the selection. Yet, repeated usage also leads to higher effect size composites and, hence, lower diversity goal values. The result of both tendencies is that the Pareto-optimal trade-offs showing higher diversity values get dominated by the corresponding trade-offs under the stage-specific predictor usage condition, and that Pareto-optimal trade-offs with very high-associated diversity values are no longer achievable.

The above explanation also suggests that combining all presently and previously administered predictors when forming predictor composites will offer better results if the contextual constraints on the weighting of predictors into predictor composites are fairly relaxed. The bottom panel of Figure 6, comparing the Pareto front for the all predictor usage condition (dotted curve) and the stage-specific usage condition (solid curve) as both obtained for the general constrained situation in which the ratio restriction on the predictor weighs is removed, corroborates this suggestion.

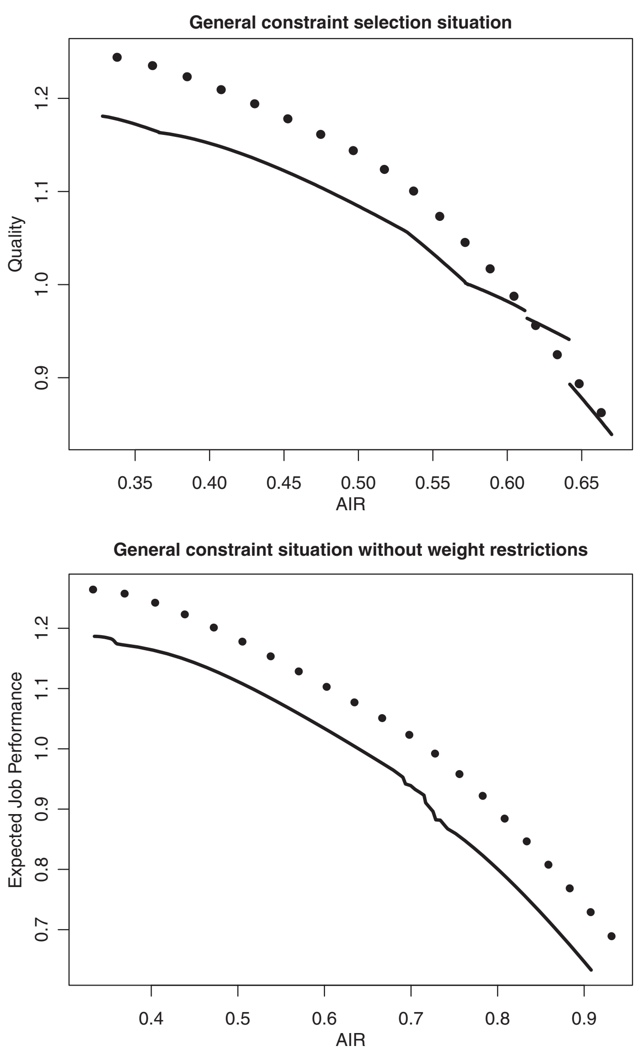

Figure 6. Impact of using only the stage-specific predictors versus using all thus far administered predictors in forming predictor composites in the general constrained selection situation (upper panel) and the general constrained situation without the ratio constraint on the predictor weights (lower panel). Both panels compare the selection quality/diversity Pareto front using only stage-specific predictors (solid curve) and using all thus far available predictors (dotted curve). AIR = adverse impact ratio.

General Discussion

Methodological and Practical Contributions

The article introduced an analytic method for designing selection systems that optimize the trade-off between competing selection outcomes. The method integrates the analytic estimation of selection outcomes in general multi-stage selection within a multi-objective optimization framework, resulting in the determination of Pareto-optimal selection systems and, hence, in the design of operational selection systems that show Pareto-optimal trade-off values for the valued selection goals. We showed how the method can assist the selection practitioner in making six major selection design decisions: (1) the predictor subset decision, (2) the selection rule decision, (3) the selection staging decision, (4) the predictor sequencing decision, (5) the predictor weighting decision, and (6) the stage retention rate decisions. Using simulation procedures, we also observed that the Pareto-optimal selection systems obtained by our method continue to perform well under realistic levels of sampling variability in the data input parameter values. This finding boosts the practical importance of the proposal because it indicates that the tool and its results are also applicable in real settings using estimates as substitute for the typically unknown actual data parameter values.

In the presentation of our method, we focused on situations where the valued selection goals are not strictly prioritized because strict goal prioritization in the sense that one first decides to maximize one of the goals and then aims for the best possible achievement on the other goal implies choosing either the quality or the diversity maximizing Pareto-optimal selection system. We also focused on the design of selection systems that show Pareto-optimal trade-offs with respect to the goals of selection quality, as gauged by the expected average job performance of the selected applicants, and work force diversity as measured by the AIR value. Yet, we emphasized that other metrics can be chosen for both selection goals and that these metrics, when monotonically related to the present ones, will lead to the same set of Pareto-optimal selection systems. We also noted several possible extensions of the method. Some of these extensions may better tune the method to certain particular characteristics of the planned selection decision. Thus, although we considered the implementation of contextual constraints related to the number, the type of usage (i.e., compensatory vs. non-compensatory), the weighting, the cost, and the sequencing of the predictors, other specifications concerning, for example, total testing time and specific logistical considerations can be taken into account as well.

Substantive Contributions

The example application illustrated the vast diversity in the trade-offs between the selection goals that can be achieved depending on the decisions made during the design of the selection system. We argued that designs that lead to a Pareto-optimal goal trade-off are definitely to be preferred over other selection designs. This argument is perfectly consistent with the current regulations in the United States and elsewhere to favor selection systems that increase selection diversity while maintaining equal levels of the selection quality objective. In fact, given these regulations, it may be quite difficult to argue the legal defensibility of selection systems that are clearly outperformed by a Pareto-optimal selection system. At the very least, it seems that in these cases, a compelling concern—unrelated to the quality and diversity objective—will be needed to justify such a non-Pareto-optimal choice. In addition, one may then wonder whether this new concern could not be entered explicitly in the selection design, either as an additional objective or as a restriction to be met by any feasible design.

Using several further applications, we showed how our method can be used to study some vexing research questions in the field of personnel selection. In particular, we addressed research questions on the importance of using different predictor weighting systems, the impact of the selection rule and the selection staging decision, and the adequacy of using composites of only the stage-specific administered predictors to make the intermediate and final-stage selection decisions. These additional studies converge to the proposal that seeking general answers or looking for generally applicable principles to resolve these issues may not be the best way to proceed because of the numerous parameters and the huge variability in situational features involved when considering general multi-stage decisions. Instead, we propose the use of our method to decide on these issues starting from the particular contextual makeup of the planned selection.

Limitations of the Method

Like any other decision aid, our method requires certain data and a number of assumptions need to be made before results can be obtained. We emphasized that these requirements are not specific to our method but are shared by all previous related work. Nevertheless, it remains important to assess whether it is realistic to assume that fairly accurate values are usually available for the required predictor, criterion, and applicant data. It is equally important to explore the limitations related to the assumptional basis of the method and to acknowledge that the results presented in the illustration and application sections are conditional on the particular chosen selection situation.

The data requirements issue has already been discussed in the literature (e.g., De Corte et al., 2006, 2007), resulting in the plea for continuing meta-analytic research on predictor and criterion characteristics and for supplementing these efforts with local validity studies. Appendix B provides a number of suggestions on how users may obtain the necessary predictor input data, but additional data (especially at the applicant rather than the incumbent population level and related to other than black–white mean predictor differences) are badly needed to ensure a wider applicability of the method. Finally, although the sensitivity analysis revealed that the results of the method generalize fairly well under realistic levels of sampling variability in the input data values, it is important to emphasize that good quality data remain essential for obtaining useful results.

As to the assumptions invoked by this and other previous related methods, it has already been noted that some of the assumptions can be relaxed quite easily (e.g., De Corte et al., 2006). However, relaxing these assumptions often requires additional, but typically unavailable, applicant and predictor data and/or leads to formidable computational challenges. Consider for example the assumption of the equality in the different applicant subpopulations of the predictor/criterion variance/covariance matrix. Although the computations in the decision tool can easily be extended to account for different variance/covariance matrices in the majority and minority populations, the current paucity of reliable results on this difference does not encourage such an extension. As a second example, one could consider invoking multi-variate distributions for the predictor and criterion variables that are more general than the present multi-normal model (e.g., elliptical distributions), but this will not only require more data on the actual distribution of predictor scores in applicant samples than is presently available but will also lead to severe computational problems.

Maintaining the assumption that the criterion and the predictors follow a multi-variate normal distribution with the same variance/covariance matrix in each applicant subpopulation (cf. Appendix A) raises some further, potentially limiting issues. First, and as observed by one of the reviewers, the assumption implies that the predictor/criterion variance/covariance matrix at the subpopulation level will differ from the corresponding matrix at the total population level when the predictors and the criterion have different means in the subpopulations (cf. Day, 1969). The implication is that using total population variance and covariance estimates as proxy for the subpopulation values, as is done in the computations of the decision aid, introduces a source of systematic error in the input data. However, the bias will often be quite small (and typically substantially smaller compared with the bias resulting from sampling variability) when the majority and minority applicant representation is substantially different and the largest predictor effect size value is less than one. For example, using Table 1 data and assuming a 12% versus 88% minority/majority representation, the application of Day's (1969) formula on the (co)variance in mixture populations (cf. Mardia, Kent, & Bibby, 1988, p.

388) shows an average absolute difference of .008 between the total population and the subpopulation variance/covariance parameter values, whereas the corresponding average absolute difference, due to sampling variability with sample sizes of 100 and 500, equals .076 and .034, respectively. Finally note that Equation 1 presented by Day (1969, p. 464) can, if required, be used to adjust the within subpopulation covariance matrices.

Second, the multi-normal assumption implies linear and, hence, monotone predictor criterion relationships. However, in certain situations other types of relationship, such as represented by minimum qualification step functions, may be more appropriate. Yet, even in these cases, the decision aid can still be applicable albeit using the selection success rate rather than the expected performance metric for gauging the selection quality objective. This switch of metric seems particularly indicated when criterion performance is evaluated as a dichotomy with success or failure corresponding to a particular minimum performance level but where it is still reasonable to posit linearity in the underlying predictor criterion relationship. Under these conditions, there is no conflict between the multi-normal assumption and the nature of the valued criterion performance, and the decision aid can be focused on the design of selection systems that offer Pareto-optimal trade-offs between selection diversity and the success rate of the selection.

Finally, one may argue against the multi-variate normal assumption because applicant groups are often more adequately perceived as non-random samples from otherwise normally distributed applicant populations, with the non-random sampling representing the result of some pre-selection screening process. If that is the case, then the situation can still be handled within the present framework provided that information is available on the relation of the pre-selection screen with the actual predictors and criterion. With this information, the initial screen can be equated to the first stage of the selection process, and our method can be applied to the thus modified selection situation.

From the previous discussion, it is clear that our method should be perceived as a decision aid and not as a device that automatically generates the right decision. When the initial application conditions in terms of the required data and assumptions cannot be reasonably met, not even after the above discusses adjustments, then the method should not be applied. Alternatively, when the application conditions are reasonably fulfilled, the method offers a unique contribution to the design of future planned selections. In the next subsection, we explore some further challenges that, when addressed properly, may make the method into an even more realistic and practically applicable tool for both applied and theoretical research.

Future Developments

Designing optimal selection systems, as is discussed here, represents one option to address the selection quality/adverse impact quandary. Other options—such as banding, the development of new low impact screening devices, or using innovative test presentation and response formats—offer other routes to achieve diversity in the work place. Also, selection is only one stage in the entire staffing process. Recently, Newman and Lyon (2009) demonstrated that the preceding recruitment stage may substantially influence the potential of any selection system to reduce adverse impact. Another aspect that can drastically affect the expected results of a planned selection system concerns applicant withdrawal during the selection process and job refusal at the end of the selection (Murphy, 1986; Tam, Murphy, & Lyall, 2004). Because recruitment efforts shape the applicant input to the selection system, whereas withdrawal and job refusal affect the output, a truly comprehensive approach should aim to account for both the recruitment and the dropout aspects in the design of optimal selection systems.

Integrating recruitment, selection, and dropout could lead to a decision aid that covers the broader staffing process. To illustrate the benefits of such an integration, consider the following example. It is a truism that dropout is more likely to happen in multi-stage selection scenarios than in single-stage selections. Yet, this impediment, which should be taken into account when comparing the potential of single- versus multi-stage selection, cannot be captured by the present proposal for designing optimal selection systems. Integrating selection and dropout also offers an opportunity to account for the fact that applicants often apply for different jobs with different organizations. Finally, an integrated model may revive the interest in applying utility theory and concepts for gauging the impact of human resource staffing practices.

Final Comment

The present proposal for designing optimal selection systems is best perceived as one alternative toward shaping selections that are likely to result in well balanced trade-offs for desired selection outcomes. It is also recognized that the conceptual basis of the method, as reflected in the notion of Pareto-optimality, is not entirely undisputed (cf. Kehoe, 2008; Potosky, Bobko, & Roth, 2008; Sackett, De Corte, & Lievens, 2008). In addition, the proposal focuses on selection decisions, although other staffing decisions—and, in particular, placement or classification decisions—may become more prominent in the near future (cf. Landy & Conte, 2007). However, compared with the present state of affairs, it offers the first systematic and integrated attempt to make the most of the available information for the purpose of optimizing the design of planned selections. Finally, given the potential of the method to study major selection research questions as well as to achieve selection designs that show substantially better trade-offs between valued goals than is otherwise possible, we encourage selection practitioners and researchers to consider applying the method whenever possible.

References

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. doi:10.1177/1094428106294734

Boggs, P. T., & Tolle, J. W. (1995). Sequential quadratic programming. *Acta Numerica, 4,* 1–51. doi:10.1017/S0962492900002518

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,* 631–657. doi:10.1137/S1052623496307510

Day, N. E. (1969). Estimating components of a mixture of normal distributions. *Biometrika, 56,* 463–474. doi:10.1093/biomet/56.3.463

De Corte, W., Lievens, F., & Sackett, P. R. (2006). Predicting adverse impact and multistage mean criterion performance in selection. *Journal of Applied Psychology, 91,* 523–537. doi:10.1037/0021-9010.91.3.523

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. doi:10.1037/0021-9010.92.5.1380

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. doi:10.1111/j.1468-2389.2008.00423.x

De Corte, W., Sackett, P. R., & Lievens, F. (2010). Selecting predictor subsets considering validity and adverse impact. *International Journal of Selection and Assessment, 18,* 260–270. doi:10.1111/j.1468-2389.2010.00509.x

Doverspike, D., Winter, J. L., Healy, M. C., & Barrett, G. V. (1996). Simulations as a method of illustrating the impact of differential weights on personnel selection outcomes. *Human Performance, 9,* 259–273. doi:10.1207/s15327043hup0903_5

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. doi:10.1037/a0013775

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker–Koopman–Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments, & Computers, 31,* 727–730. doi:10.3758/BF03200754

Johnson, J. W., Abrahams, N., & Held, J. D. (2004). *A procedure for selecting predictors considering validity and adverse impact.* Poster presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

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. doi:10.1111/j.1468-2389.2008.00424.x

Landy, F. J., & Conte, J. M. (2007). *Work in the 21st century: An introduction to industrial and organizational psychology* (2nd ed.). Malden, MA: Blackwell.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1988). *Multivariate analysis.* New York, NY: Academic Press.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85,* 812–821. doi:10.1037/0021-9010.85.5.812

Messac, A., Ismail-Yahaya, A., & Mattson, C. A. (2003). The normalized normal constraint method for generating the Pareto frontier. *Structural and Multidisciplinary Optimization, 25,* 86–98. doi:10.1007/s00158-002-0276-1

Murphy, K. R. (1986). When your top choice turns you down—Effect of rejected offers on the utility of selection tests. *Psychological Bulletin, 99,* 133–138. doi:10.1037/0033-2909.99.1.133

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. doi:10.1037/a0013472

Ones, D. S., & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83,* 35–42. doi:10.1037/0021-9010.83.1.35

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679–703. doi:10.1037/0021-9010.78.4.679

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. doi:10.1111/j.1744-6570.2008.00109.x

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,* 304–315. doi:10.1111/j.1468-2389.2005.00327.x

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. doi:10.1111/j.1468-2389.2008.00425.x

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. doi:10.1111/j.1468-2389.2008.00426.x

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49,* 549–572. doi:10.1111/j.1744-6570.1996.tb01584.x

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49,* 787–829. doi:10.1111/j.1744-6570.1996.tb02450.x

Schmidt, F. L., & Hunter, J. (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. doi:10.1037/0033-2909.124.2.262

Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology, 84,* 50–57. doi:10.1037/0021-9010.84.1.50

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. doi:10.1111/j.1744-6570.2004.00010.x

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., Jr., & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17,* 71–93. doi:10.1207/S15327043HUP1701_4

(*Appendices follow*)

Appendix A

Obtaining Pareto-Optimal Selection Systems

This appendix details the method used to compute the Pareto-optimal selection systems and the corresponding Pareto-optimal trade-offs. For reasons of clarity, the description focuses on the case where both the selection quality and diversity objective are of importance. The method proceeds in three main steps: identification of the feasible selection scenarios, computation of the Pareto-optimal front and the corresponding Pareto-optimal selection systems for each feasible scenario, and aggregation of the scenario specific solutions across all feasible scenarios to obtain the global Pareto-optimal front.

The set of all feasible selection scenarios is generated by integrating a combinatorial routine that enumerates all possible scenarios with a filter procedure that retains only the scenarios that meet the structural contextual requirements in terms of the number and the staging of the predictors. Scenarios that involve non-compensatory predictors are further transformed to the appropriate multiple hurdle scenario.

Next, the normal boundary intersection procedure (Das & Dennis, 1998) is applied to each of the feasible scenarios in succession. For each scenario, this implementation proceeds in two stages that both imply solving a number of constrained non-linear programming problems using a sequential quadratic programming scheme (cf. Boggs & Tolle, 1995). The two non-linear programs of the first stage compute the selection systems that optimize the selection quality and diversity goal, respectively. The objective function of the programs corresponds to the selection goal that is to be maximized, whereas the problem variables are the set **B** of weights with which the predictors are combined to the stage specific predictor composites and the set **y** of cutoff scores to be used with respect to the composites in the different selection stages. By default, the predictor weights obey a non-negativity constraint, but the constraint can be replaced by other restrictions (e.g., constraints on the ratio of pairs of predictor weights to ensure that all the predictors that are administered in a stage effectively contribute to the composite formation). The programs also share the set of constraints that translate the non-structural contextual requirements (e.g., the time, cost, predictor weighting, and retention rate constraints) formulated in the preparatory analysis phase. At the solution, the first program results in the maximum possible value for the selection quality, denoted as q_{max}, and the corresponding value a_q for the diversity objective, whereas the solution of the second program returns the maximum possible adverse impact ratio value, a_{max}, and the corresponding quality value q_a. The first pair of values (q_{max}, a_q) is referred to in the text as the *quality maximizing Pareto-optimal trade-off*, and the second pair (q_a, a_{max}) is called the *diversity maximizing Pareto-optimal trade-off*.

In the second stage of the normal boundary intersection method, the quality and diversity maximizing trade-offs are used to formulate and solve a series of new constrained non-linear programs, with each program resulting in a new Pareto-optimal system that shows a more balanced quality/diversity trade-off. The problem variables of these (maximization) programs are **B** and **y**, augmented by a new problem variable r, which also constitutes the objective function of these programs. Except for one addition, these programs have the same constraints as the programs solved in the first stage. The added equality constraint is $\boldsymbol{\Phi\beta} - r\boldsymbol{\Phi 1} = \mathbf{v}_z$, where $\boldsymbol{\Phi}$ is a 2 × 2 payoff matrix with rows (0, $q_a - q_{max}$) and ($a_q - a_{max}$, 0); $\boldsymbol{\beta} = (\beta, 1\text{-}\beta)'$; and $\mathbf{v}_z = \mathbf{t}_z - \mathbf{t}_0$, with $\mathbf{t}_0 = (q_{max}, a_{max})'$ and $\mathbf{t}_z = (q_z, a_z)'$, where q_z and a_z are the values of the selection quality and the selection diversity objective associated with the values **z** for the weights **B** and the cutoffs **y** (cf. De Corte, Lievens, & Sackett, 2007). So, the programs of the second stage all share the same makeup, but they differ in the value of β, which takes equally spaced values between 0 and 1. Also note that the representation of the Pareto-optimal front can be made as detailed as desired by solving the Stage 2 non-linear programs for an increasing number of equally spaced β values.

Solving the above detailed nonlinear programs requires the repeated computation of the selection outcome values for varying sets of values for the program problem variables. To perform these computations, it is assumed that the predictor and criterion variables have a multi-variate normal distribution with the same variance/covariance matrix but a different mean vector in each of the applicant subpopulations. In particular, and without loss of generality, it is assumed that the predictor and criterion variables have a standard multi-variate normal distribution in the majority population. As shown by De Corte, Lievens, and Sackett (2006), the assumption permits deriving analytic expressions for the selection outcome values within general multi-stage selection settings, and we use these expressions to perform the calculations required by the implementation of the normal-boundary method.

In the third and final step of our method, the Pareto filter algorithm described by Messac, Ismail-Yahaya, and Mattson (2003, p. 93) is applied to the entire collection of Pareto-optimal trade-offs as obtained over the different feasible selection scenarios. This filter algorithm essentially locates and retains the Pareto-optimal trade-offs within the entire collection that are not dominated by any other trade-off of the collection.

The above described method also applies to situations where one intends designing unit weighted or regression weighted Pareto-optimal selection systems. However, in that case, the set of problem variables corresponding to the different non-linear programming problems no longer comprises the set **B** of predictor weights, and the computation of the selection outcomes is performed using either unit or regression weights for the predictors in the formation of the predictor composites.

(*Appendices continue*)

Appendix B

Applying the Method

This appendix addresses the two major issues regarding the applied use of the proposed method: obtaining the required input data and using the software to implement the method.

Obtaining the Input Data

As we have noted, the application of the proposed method requires estimates for the values of the predictor-criterion correlations, the interpredictor correlations, and the effect sizes regarding subgroup differences. We address in turn each of the three needed sets of estimates.

First, the method requires estimates of the validity of each predictor, at level of the applicant pool. In essence, this means that one needs a range-restriction corrected validity estimate for each predictor. Note that to get to the point of asking questions about *how* to use predictors (which is the purpose of the method proposed here), the user must have already decided that the predictor meets some validity threshold. That decision may have been based on a local validity study or on some form of generalized validity evidence (e.g., transporting validity from a closely related setting, or reliance on meta-analytic findings). Thus, a validity estimate is already in hand. It is possible, though, that the validity estimate is based on a selected sample (e.g., via a concurrent validity study). If that is the case, the user needs to estimate the degree of range restriction in the sample used as the basis for the validity estimate, and apply a correction formula. Sometimes the data needed for such a correction (i.e., the predictor standard deviation in the applicant pool and in the selected sample) are readily at hand, as in the case where the measure has been administered to applicant samples either for research purposes or as part of prior ongoing operational use. In other settings, the needed data may not be readily available, and professional judgment may be the basis for an estimate. For example, one might conclude that there is no evidence that existing selection system in the organization screened on the construct of interest and, thus, judge that the validity estimate obtained from an incumbent sample can serve as a good estimate of validity in the applicant pool.

Second, the method requires estimates of the correlations among all predictors, again, at the level of the applicant pool. In general, it is easier to obtain these data at the applicant level than it is to obtain validity data, as in many cases it is possible to administer experimental predictors to applicants prior to operational use, or to obtain these data from operational applicant settings if a set of predictors is in use and the goal of applying the current article's approach is to re-evaluate how the predictors are used. In other settings, though, the only local data may be from a study with incumbents, in which the same need for range restriction exists as in the discussion of validity above. Also, in the case of some commonly used predictors, a broader literature, and perhaps meta-analytic syntheses, exists regarding interpredictor correlations.

Third, the method requires estimates of mean subgroup differences at the applicant pool level for each predictor. Issues regarding the availability of these data generally parallel the discussion above regarding interpredictor correlations. (i.e., the same local data that are used to calculate interpredictor correlations can be used to calculate mean group differences). One additional complication is that in the case of some subgroups, the number of group members in the applicant pool may be too small to yield credible estimates. In such cases, reliance on effects reported in the literature may be preferable.

Obviously, there will be settings where all the needed estimates will be relatively readily available, and others where the estimation will be more difficult and will rely more heavily on professional judgment. However, obtaining the needed estimates strikes us as a manageable task in many, if not most, settings. We reject the notion that the methods we develop in this article are only of academic interest. Instead, we note that our approach formalizes a set of considerations that are, at least implicitly, involved in any selection system design where both criterion performance and diversity are outcomes of interest. Both the data needs of the present methods and the set of six decisions in selection system design (i.e., predictor subset selection, selection rule, selection staging, predictor sequencing, predictor weighting, and stage retention) are implicit considerations that our approach makes explicit. Our approach makes no exceptional demands: It simply requires fuller articulation of the issues in selection system design.

Finally, we acknowledge that there may be instances where most of the needed data will be available, but one or two pieces of information cannot be estimated with a level of confidence that makes the user comfortable. Here, we suggest applying our method multiple times, imputing one's best estimate bracketed by others (e.g., "worst case scenarios") to determine the degree to which conclusions about optimal systems are affected by this uncertainty.

Getting and Implementing the Software

A demo executable version of the decision aid can be downloaded from http://users.ugent.be/~wdecorte/software.html. To run the executable, a 32 bits Windows operating system (XP and beyond) is required. The site also provides a link to a manual that details the usage of the software. The tutorial and the demo executable of the decision aid focus on the situation where the practitioner intends designing selection systems that offer a Pareto-optimal trade-off between the goals of selection quality and diversity and where the applicant pool comes from a mixture of two populations: the majority population and a single minority population.

(*Appendices continue*)

Before implementing the decision aid, the user must first perform the subtasks related to (a) the inventory of the available predictor battery and the characterization of the current applicant pool and (b) the specification of the contextual constraints. The articulation of the selection goals, as required by the first subtask, is already implemented in the decision aid: The quality goal is equated to the expected job performance of the selected applicants, and the diversity goal is translated by means of the adverse impact ratio. The characterization of the current applicant pool reduces to the specification of the proportion of majority and minority applicant candidates in the current or the expected applicant pool, whereas the data inventory concerning the available predictor battery has been discussed above.

The current implementation of the decision tool supports specification of the contextual constraints in terms of (a) the total number of predictors and selection stages, (b) the number and the staging of the predictors, (c) the usage (compensatory or noncompensatory) of the predictors, (d) the maximum weight ratio when forming predictor composites, (e) the minimum and maximum permissible retention rates in the intermediate selection stages, and (f) the acceptable total predictor cost value. In addition it is possible to require either unit or regression weighting when forming predictor composites or to implement a screening strategy in which at each stage all thus far available predictor information is used when forming predictor composites.

The manual that accompanies the decision aid software details how the information resulting from the preparatory stage can be coded to the required input format and how the program can be executed. After execution, the decision aid returns an output file that contains the information as described in the article.

Received January 18, 2010
Revision received January 20, 2011
Accepted February 15, 2011 ■