9-2008

Validity and adverse impact potential of predictor composite formation

Wilfried DE CORTE
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Paul R. SACKETT
University of Minnesota - Twin Cities

DOI: https://doi.org/10.1111/j.1468-2389.2008.00423.x

Validity and Adverse Impact Potential of Predictor Composite Formation

Wilfried De Corte, Department of Data Analysis, Ghent University, Ghent, Belgium. wilfried.decorte@ugent.be

Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Ghent, Belgium

Paul R. Sackett, University of Minnesota, Twin Cities Campus, Minneapolis, MN, USA

Published in International Journal of Selection and Assessment, 2008 September, 16 (3), 183-194.

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

Submitted version

Abstract

Previous research on the validity and adverse impact (AI) of predictor composite formation focused on the merits of regression-based or ad hoc composites. We argue for a broader focus. Ad hoc chosen composites are usually not Pareto-optimal, whereas the regression-based composite represents only one element from the total set of Pareto-optimal composites and can, therefore, provide only limited information on the potential for validity and AI reduction of forming predictor composites when both validity and AI are of concern. In that case, other Pareto-optimal composites may provide a better benchmark to decide on the merits of the predictor composite formation. We summarize a method to determine the set of Pareto-optimal composites and apply the method to a representative collection of selection predictors. The application shows that the assessment of the AI and validity of predictor composite formation can differ substantially from the one arrived at when considering only regression-based composites.

1. Introduction

In a recent paper, Potosky, Bobko, and Roth (2005) present an updated, meta-analytically derived matrix of the validity and effect size (i.e., the standardized average group difference) of four commonly administered selection predictors. The four predictors are cognitive ability (CA) and three non-cognitive predictors: conscientiousness (CO), structured interview (SI) and biodata (BI). They subsequently use the updated estimates to study the fairly widespread belief that selection composites of CA and a non-cognitive predictor will typically reduce adverse impact (AI) as compared with when CA is used alone. At the end of their investigation, they conclude that the strategy of adding a non-cognitive measure to a CA test often results in only relatively modest decreases in AI (cf. Potosky *et al.* (2005) p. 311).

Although this conclusion is based on probably the best available data on the validity and effect size of currently employed predictors, we show below that it requires some amendment when considering the full potential of forming predictor composites which both increase the validity and reduce the AI of selection decisions. Our argument proceeds in two stages. First, we summarize a recently developed method for determining the AI and validity potential of composite predictors relative to the AI and validity of a single cognitive predictor. Second, we apply this method to the Potosky *et al.* (2005) data and obtain results that indicate a substantially better performance of the considered composites than the one suggested by Potosky *et al.*

2. Determination of the validity and AI potential of predictor composites

Research on the effects of forming composites of various predictors on validity and AI has typically focused on regression weighting, unit weighting, or various *ad hoc* weighting schemes (Sackett & Ellingson, 1997; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). *Regression weighting* is commonly examined as it is the method that maximizes the performance of those selected, given the set of available predictors. *Unit weighting* is commonly examined due to the simplicity of the approach, the recognition that observed regression weights are subject to sampling error, and the finding that unit weights often perform surprisingly well relative to regression weights (Dana & Dawes, 2004; Bobko, Roth, & Buster, 2007). *Ad hoc weighting* schemes are commonly examined in pursuit of alternative weightings that fare better in terms of AI at a relatively small decrement in validity (e.g., Doverspike, Winter, Healy, & Barrett, 1996; Hattrup & Rock, 2002). Such strategies are exemplified by technical reports that typically present a trial and error examination of a series of alternative models, which use varying combinations of available predictors, weighted in differing ways.

A recent paper by De Corte, Lievens, and Sackett (2007) offered a technical and formal solution to questions about how one might best balance productivity (i.e., high validity) and diversity (i.e., low AI) objectives. This paper presented the notion of Pareto-optimal trade-offs between the two outcomes. Given a set of predictors, there are an infinite number of possible weighting schemes that could be applied in forming predictor composites. A Pareto-optimal trade-off is a weighting scheme for which one outcome cannot be improved without harm to the other outcome. For example, there may be multiple weighting schemes, which would result in a given level of validity; of these schemes, the Pareto-optimal one is the set of weights, which result in the highest adverse impact ratio (AIR). Similarly, there may be multiple weighting schemes, which would result in a given AIR; the Pareto-optimal one is the set of weights, which result in the highest level of validity. Therefore, Pareto-optimal composites offer optimal trade-offs between the AI and the validity objective, and the entire collection of these Pareto-optimal trade-offs is usually referred to as the Pareto-optimal trade-off curve or function (Keeney & Raiffa, 1993; Pareto, 1906).

The definition of the set of Pareto-optimal composites implies that the *regression-based composite* is one particular element of the set. As regression-based weights maximize the validity of the resulting composite, no other weighing of the predictors can outperform this composite in terms of the validity criterion. The *minimal* *impact composite*, defined as the composite with the highest possible AIR value, is another example of the set. Under the common condition of all positive predictor effect sizes, the regression-based and the minimum impact composite are the boundary points of the Pareto-optimal set, with all the other Pareto-optimal composites showing more balanced trade-offs between validity and AI. More specifically, these intermediate composites are all characterized by a smaller validity than the regression-based composite and they all show a smaller value for the AIR than the minimum impact composite.

De Corte *et al.* (2007) provide details on the multi-criteria optimization procedure used to identify the set of Pareto-optimal composites, given a set of predictors with given validity, intercorrelations and subgroup differences, and specifying a selection ratio. They present the procedure as a decision tool to assist selection practitioners when planning future selection systems in situations where both AI and selection quality are of concern. There will be many instances where it is advantageous, or even necessary, to determine predictor weights in advance of obtaining predictor data from applicants. For example, in some settings organizations are required by statute or policy to reveal the weights given to the components of a selection system to applicants before testing. Investigating weighting schemes *a priori* may be legally more defensible than waiting until after predictor data have been gathered, an approach that also underlies several earlier proposals to address the expected impact of differential predictor weights on personnel selection outcomes (e.g., Doverspike *et al.*, 1996).

The multi-criteria optimization procedure of De Corte *et al.* (2007) proceeds in two stages. In the first stage, two constrained non-linear programming problems are solved to obtain the predictor weighing schemes that result in the maximum possible value for the selection quality objective (i.e., the composite validity) and the maximum possible value of the AI objective, respectively. These weighing schemes characterize two optimal trade-off points, where the first optimal trade-off (i.e., the trade-off associated with the regression-based composite) corresponds to the situation where only the validity objective is judged to be of importance, and the second optimal trade-off (associated with the minimal impact composite) represents the situation where only the AI objective is of concern. These two validity/AI trade-offs permit the determination of a payoff matrix that is subsequently used in the formulation of a series of non-linear programs that are solved in the second stage of the method. These second stage non-linear programs result in new optimal trade-offs between composite validity and composite AI, each of which corresponds to a particular predictor-weighing scheme and a particular valuation of the two selection objectives. Together with the earlier obtained trade-offs, these new trade-off points and their associated predictor weighing schemes (each of which characterizes a Pareto-optimal composite) provide an evenly spaced, representative characterization of the entire set of such Pareto-optimal composites.

The results of the procedure can be expressed in tabular or graphical form. Figure 1illustrates the graphical outcome of the technique; it presents the Pareto-optimal trade-off curve for a composite of CA and a SI, based on values from Potosky *et al.* (2005) (cf. Table 1). The figure shows the optimal levels of AIR achievable at each level of validity, or, equivalently, the optimal level of validity achievable at each level of AIR. Table 2 shows the tabular presentation as it further details a selected number of optimal trade-offs. For each selected trade-off (cf. the numbered trade-off points on Figure 1), the table summarizes the validity and AI ratio value as well as the weighting (with weights scaled to have unit sum) of the predictors that characterize the corresponding optimal composite.

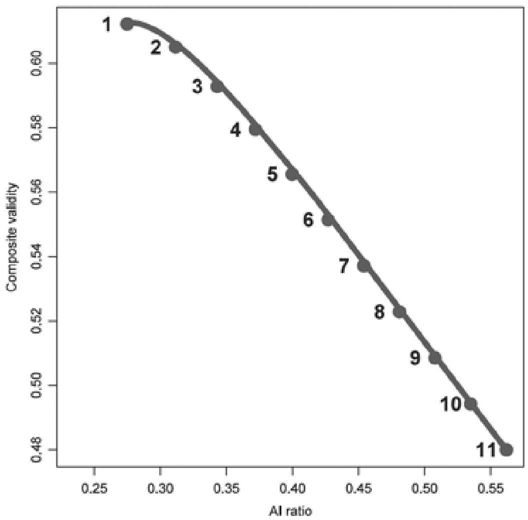

Figure 1. Pareto-optimal validity–adverse impact ratio trade-off curve for a selection with selection rate of .10 using a composite of cognitive ability and a structured interview, as based on values from Potosky *et al.* (2005) (cf. Table 1).

Table 1. Predictor effect sizes, validities and intercorrelations

Variable	Effect size[a]	Validity	Intercorrelation matrix			
			1	2	3	4
Predictors						
1. Cognitive ability (CA)	.72	.51				
2. Structured interview (SI)	.31	.48	.31			
3. Conscientiousness (CO)	.06	.22	.03	.26		
4. Biodata (BI)	.57	.32	.37	.17	.31	

Note: [a]Effect sizes are relative to the minority applicant population.

Table 2. Selected Pareto-optimal trade-off composites of cognitive ability (CA) and structured interview (SI)

Point	Validity	AI ratio	Predictor weights	
			CA	SI
1	.61	.27	.53	.47
2	.61	.31	.42	.58
3	.59	.34	.35	.65
4	.58	.37	.30	.70
5	.57	.40	.25	.75
6	.55	.43	.20	.80
7	.54	.45	.16	.84
8	.52	.48	.12	.88
9	.51	.51	.08	.92
10	.49	.53	.04	.96
11	.48	.56	.00	1.00

De Corte *et al.* (2007) present the approach as a method of choosing among differing weighting schemes, given a set of predictors. For example, one might take the Pareto-optimal solution that maximizes validity as a starting point and ask ‘what improvement in the AIR would be achievable if one is willing to accept an *X*% reduction in validity?’ (where *X*reflects a judgment about the reduction in validity that a researcher may be willing to accept in the interest of an increase in diversity, such as 1%, 5%, or 10%). We emphasize the phrase ‘is willing to accept’ in the above sentence: the approach does not specify any particular trade-off that one should accept.

The present paper adds one additional feature to this formulation in response to Potosky *et al.*'s (2005) conclusion that supplementing a CA measure with additional predictors commonly produces only a modest reduction in AI. To introduce this feature we refer to Figure 2 which adds one piece of information to Figure 1, namely, the level of validity and the AIR (i.e., the validity–AIR trade-off) obtained by using CA alone as a predictor, that trade-off is represented by the diamond shape symbol in the figure. Figure 2 allows us to make our main point, namely, that Potosky *et al.* compare the use of CA alone with only one alternative Pareto-optimal solution, namely, the regression-weighted composite of the predictors. As Figure 2 shows, use of the regression-weighted composite of CA and the SI (the uppermost point on the Pareto-optimal trade-off curve) results in only a very small change in the AIR: from .23 to .27. It also produces an improvement in validity from .51 to .61.

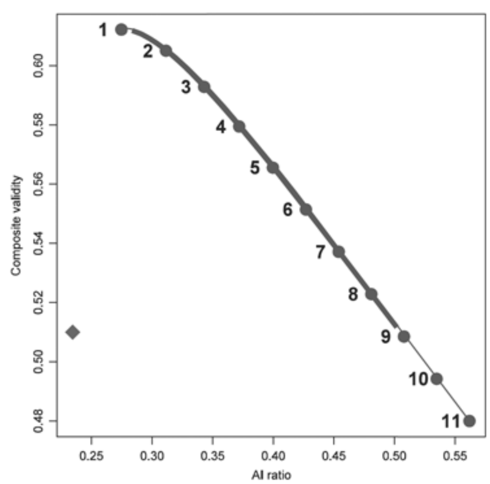

Figure 2. Pareto-optimal validity–adverse impact ratio trade-offs achievable using a composite of cognitive ability and a structured interview compared with using cognitive ability alone. The bold part of the Pareto-optimal curve summarizes the trade-offs that dominate the cognitive ability validity and AIR value.

Although using the regression-weighted composite has received considerable focus, as it maximizes the increment in validity from adding an additional predictor (e.g., Schmitt *et al.*, 1997), we propose that a broader perspective on the impact of adding a new predictor on validity and AI reduction is possible by considering not only the regression-weighted composite but other Pareto-optimal composites as well. It is certainly true that the regression-weighted composite is the only viable choice when only the validity potential of adding a non-cognitive predictor to CA is of interest because this composite maximizes validity and views any accompanying increase in the AIR as an incidental benefit. However, several authors (Hoffman & Thornton, 1997; Hough, Oswald, & Ployhart, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001) have warned against the one-sided quest to maximize validity as many organizations seek a balance between diversity concerns and performance outcomes. In addition, the common argument for the use of additional predictors to supplement CA is that the use of such predictors has the potential to both increase validity and reduce AI. If one is investigating additional predictors with the joint goals of increasing validity and reducing AI, one might consider a broader range of solutions, which differ in their relative emphasis on increasing validity and reducing AI. To this end, the method proposed by De Corte *et al.* (2007) is of particular relevance because it enables selection system developers to find these other Pareto-optimal composites. Thus, the bold portion of the Pareto-optimal trade-off curve in Figure 2 identifies all Pareto-optimal composites, which increase both validity and the AIR. This figure makes clear that a much greater improvement in AIR is achievable, along with an improvement in validity, if one is willing to consider Pareto-optimal solutions other than the one that maximizes validity. For example, point number six in the figure represents a set of weights applied to the cognitive and interview predictors (i.e., .20 and .80, respectively, cf. Table 2), which produces an AIR of .43 (relative to the value of .23 with ability alone) and validity of .55 (relative to the value of .51 with ability alone). This point reflects accepting a 10% reduction in validity from the maximum value of .61 obtained with these two predictors in return for an improvement of the AIR from .23 to .43: an 87% improvement. Whether one finds this trade-off acceptable reflects the relative value one applies to validity and to AI reduction. But the central point is that a considerably greater improvement in AIR, accompanied by an increase in validity, is possible if one assigns values to validity and AIR other than the assignment of 100% weight to validity and 0% to AIR implicitly chosen by Potosky *et al.*

Possible reactions to the approach offered here include questions about whether it is permissible to deviate from a validity maximization strategy, and whether the Civil Rights Act of 1991 precludes any selection strategy that takes AI into account when weighting predictors. Regarding the first issue, there is no general requirement to maximize validity; in fact the use of methods that depart from validity maximization is routine. Unit weights are often used for administrative ease; score bands (e.g., 'green–yellow–red' or 'pass–fail') are commonly used to simplify decision making; shorter forms of tests are commonly used to reduce costs and testing time. What is restricted by the US Civil Rights Act of 1991 is treating scores differently by subgroup. The key point is that our approach does not involve such differential treatment. All candidates are treated the same: any decision about the predictor weights applies to all of the candidates. The procedure simply includes work force diversity as an additional objective to be met by the selection system. Note that the approach does not tell the selection system designer what weights should be used. Rather, it provides information as to relative gains and losses in terms of validity and AI if differing weights are chosen, and it is a matter of values as to whether a given reduction in validity (i.e., 1% or 5%) would be deemed acceptable for a given reduction in AI.

3. Application

3.1. Context and purpose

This section reports two applications that further illustrate the potential of composite formation to improve validity and reduce AI. Both applications start from the meta-analytic matrix with estimates of the validities, effect sizes and intercorrelations of four predictors (i.e., CA, SI, CO and BI) as presented by Potosky *et al.* (2005). Table 1 summarizes these estimates. To determine the AI values, estimates for the proportion of majority and minority group applicants are required as well. For reasons of comparability with the Potosky *et al.*study, proportions equal to .881 and .119 are used throughout.

Above, we graphically examined the effects on validity and AI of forming composites of CA and the interview with a selection ratio of .10. To ascertain that our findings generalize over different values for the overall selection rate, *s*, the above analyses were repeated for *s*equal to .2, …, .9. Table 3 provides a general overview of the AIR and validity potential of selected CA and SI composites for these varying levels of the overall selection ratio. For each selection ratio the table lists the validity value, *v*, and the AIR value, *a*, of (a) the CA predictor, (b) the regression-based composite of CA and SI (i.e., the Pareto-optimal composite that has 100% maximum validity) and (c) the Pareto-optimal composites that have validity equal to 99%, 97.5%, 95%, 90% and 80% of the maximum validity associated with the regression-based composite. The predictor weights (scaled to unit sum) that characterize these Pareto-optimal composites are indicated as well. Also, to obtain a more intuitive index of the effects on diversity of the composite formation, the AIR values are further converted to percent representation of selected minority applicants relative to the representation (set to 100%) under selection using only the CA predictor. This percent minority representation (PMR) index is, henceforth, denoted as *f*. Finally, the number *n* (mentioned between brackets for the CA only composite) corresponds to the expected number of selected minority applicants when selecting from a pool of 500 applicants, whereas the number *i* (given between brackets for the Pareto-optimal composites) shows the expected gain in the number of selected minority applicants when using the composite instead of the CA only composite. Although our choice of 500 applicants in the initial hiring pool may not be typical for many selection applications, note that the expected values of the validity coefficient and the AI ratio are not dependent on sample size. We chose a relatively large applicant pool because it often happens that the same predictors are used repeatedly over a period of time such that the total applicant pool over the different selections may become quite large. Also, for small-sized selections the numbers *n* and *i* are less well suited to evaluate the effects on diversity of the composite formation because under the present 88–12% representation of the majority–minority group and for realistic selection rates the number of minority hires will then remain very small, whatever the AIR of the selection.

Table 3. AI potential of CA and SI composites at selected validity levels

SR	CA only 1.00[a] CA + .00 SI			Regression composite .53 CA + .47 SI			99% validity .43 CA + .57 SI			97.5% validity .37 CA + .63 SI		
s	*v*	*a*	*f (n)*	*v*	*a*	*f (i)*	*v*	*a*	*f (i)*	*v*	*a*	*f (i)*
.1	.51	.23	100.0 (2)	.61	.27	116.5 (+0)	.61	.31	130.2 (+0)	.60	.33	140.4 (+0)
.2	.51	.31	100.0 (4)	.61	.35	113.3 (+1)	.61	.38	124.1 (+1)	.60	.41	131.9 (+1)
.3	.51	.37	100.0 (7)	.61	.41	111.2 (+1)	.61	.45	120.0 (+2)	.60	.47	126.4 (+2)
.4	.51	.43	100.0 (11)	.61	.47	109.5 (+1)	.61	.51	116.8 (+2)	.60	.53	122.1 (+2)
.5	.51	.49	100.0 (16)	.61	.53	107.9 (+1)	.61	.57	114.0 (+2)	.60	.59	118.3 (+2)
.6	.51	.56	100.0 (21)	.61	.60	106.5 (+1)	.61	.63	111.5 (+2)	.60	.65	114.9 (+3)
.7	.51	.63	100.0 (27)	.61	.66	105.2 (+2)	.61	.69	109.1 (+3)	.60	.71	111.7 (+4)
.8	.51	.71	100.0 (35)	.61	.74	103.8 (+1)	.61	.76	106.6 (+2)	.60	.78	108.5 (+3)
.9	.51	.82	100.0 (45)	.61	.84	102.2 (+1)	.61	.85	103.9 (+2)	.60	.86	105.0 (+2)

SR	95% validity .31 CA + .69 SI			90% validity .20 CA + .80 SI			80% validity .03 CA + .97 SI		
s	*v*	*a*	*f (i)*	*v*	*a*	*f (i)*	*v*	*a*	*f (i)*
.1	.58	.37	154.0 (+0)	.55	.43	177.9 (+1)	.49	.54	222.7 (+1)
.2	.58	.44	142.3 (+2)	.55	.50	159.9 (+2)	.49	.61	191.7 (+4)
.3	.58	.51	134.7 (+3)	.55	.56	148.6 (+4)	.49	.66	173.0 (+5)
.4	.58	.56	128.8 (+3)	.55	.62	140.0 (+4)	.49	.71	159.1 (+6)
.5	.58	.62	123.8 (+3)	.55	.67	132.8 (+5)	.49	.75	147.9 (+7)
.6	.58	.67	119.3 (+4)	.55	.72	126.4 (+6)	.49	.79	138.1 (+8)
.7	.58	.73	115.1 (+5)	.55	.77	120.5 (+6)	.49	.83	129.2 (+8)
.8	.58	.80	110.9 (+4)	.55	.83	114.7 (+5)	.49	.88	120.6 (+7)
.9	.58	.88	106.3 (+3)	.55	.90	108.5 (+4)	.49	.93	111.8 (+5)

Note: [a]Weights are scaled to have unit sum.

Notwithstanding the above restriction, the numbers *n* and *i* may, as suggested by a reviewer, provide a more down-to-earth way of representing the expected increase in minority hires as compared with the PMR index *f* which may somewhat dramatize small improvements in the number of minority hires. Thus, whereas the PMR index value associated with using, for example, the 90% maximum validity composite with a selection rate of .2 equals 159.9 (thereby suggesting a substantial increase in the expected number of minority hires), actually only two additional minority hires are expected when using this composite when selecting from a sample of 500 applicants that is composed of 88.1% majority and 11.9% minority applicants. In that case, a selection with a .2 selection rate using the CA only composite is expected to lead to four minority hires (cf. the value of *n* in the second row of this scenario) whereas six minority applicants are expected to be hired when the 90% maximum validity composite is used.

Despite the somewhat sobering figures obtained for the gain in minority hires when selecting from small to medium-sized applicant samples, the results of Table 3 confirm that forming composites of the CA and the SI predictor has a noticeable validity improvement and a substantial AI reduction potential relative to using only the CA predictor. The results also show that the evaluation of the merits of composite formation when based on only the performance of the regression-based composite can seriously underestimate the potential for AIR improvement. Thus, it is found that accepting, for example, a decrease of 2.5% in the validity potential of the Pareto-optimal composite as compared with the regression-based composite leads to a substantial increase in the PMR value. For example, with a selection rate of .2 it is shown that if, say, 100 minority applicants would be selected when using only the CA predictor, then 113.3 minority hires are expected by using the regression-based composite, whereas the 97.5% maximum validity Pareto-optimal composite is estimated to generate 131.9 minority hires. Also, over the entire range of selection ratios, the increase in minority hires (compared with the hiring under CA selection) when using the 95% maximum validity Pareto-optimal composite is expected to be at least three times as large as the corresponding increase associated with the usage of the regression-based composite.

In summary, the analysis of the CA and SI composite formation indicates that balanced Pareto-optimal composites, showing decreases of 1–5% from the maximum validity, offer a substantial gain in minority representation within the selected applicants as compared with the situation where the regression-based composite is used. We suggest that other Pareto-optimal composites than the regression-based composite also merit consideration when both validity and AI are of concern.

3.2. AI and validity potential of general composite formation

So far, all the findings relate to composites of CA and one particular non-cognitive predictor, namely, the SI. However, previous researchers noted that the improvement potential of composite formation depends on certain features of the added, non-cognitive predictor (e.g., Potosky *et al.*, 2005; Sackett & Ellingson, 1997). Thus, it was observed that (all else equal) the validity improvement potential increases for higher validity levels of the non-cognitive predictor and for lower levels of correlation between the non-cognitive and the cognitive predictor. In turn, the AI reduction potential increases for lower effect size values of the non-cognitive predictor and for higher levels of correlation between the non-cognitive and the cognitive predictor. Because these tendencies are also implied by the formulae that capture the AI and the validity of predictor composites, it may be expected that the present method also shows that the merits of composite formation vary with the level of validity, effect size and correlation with CA of the added non-cognitive predictor. To verify this, we analyzed the potential merits of composites formed from CA with either CO or BI as these predictors have a different validity, effect size and CA intercorrelation pattern. Also, to provide an overview of the validity and AI potential of more general composites, we additionally studied composite formation from CA with each pair of non-cognitive predictors and from CA with all these three predictors. This requires estimates of the intercorrelations between the non-cognitive predictors that are not provided in the Potosky *et al.* (2005) study. Given that the previous analyses (as well as those of Potosky *et al.*) are all based on corrected values for the predictor effect sizes and validities, only corrected estimates of the required intercorrelations can be used, which precluded borrowing the values reported by, for example, Bobko, Roth, and Potosky (1999) or Schmitt *et al.* (1997). A literature search, using among others the Web of Science and the PsycINFO databases, converged on the corrected non-cognitive predictor intercorrelation values reported in Table 1. In particular, we used the value of .26 for the corrected correlation between CO and SI as given by Cortina, Goldstein, Payne, Davison, and Gilliland (2000), the value of .31 for the BI–CO correlation reported in Ployhart, Weekley, Holtz, and Kemp (2003) for a sample of 2544 applicants, and the value of .17 that Dalessio and Silverhart (1994) mention as the corrected estimate for the BI–SI correlation.

The results of the analyses are reported in Figure 3 and in Table 4. In particular, Figure 3 depicts for a selection ratio of .10 the curve of all Pareto-optimal AI and validity trade-offs and the subset of dominating Pareto-optimal trade-offs (cf. the bold part of the Pareto curves where the composite outperforms CA alone) associated with each studied composite. From the upper right panel of the figure it can be seen that the AI–validity trade-off associated with the CA predictor is part of the Pareto-optimal trade-off curve when forming composites from CA with the BI predictor. As a consequence, it is clear that adding BI does not result in a solution that improves both validity and AIR. Also, adding the BI predictor to composites of CA with the other non-cognitive predictors hardly changes the Pareto-optimal trade-off curve of the composite. Finally, the visual inspection of the different Pareto-optimal curves suggests that composite formation from CA, SI and CO and from CA, SI, CO and BI offers the best overall AIR and validity potential.

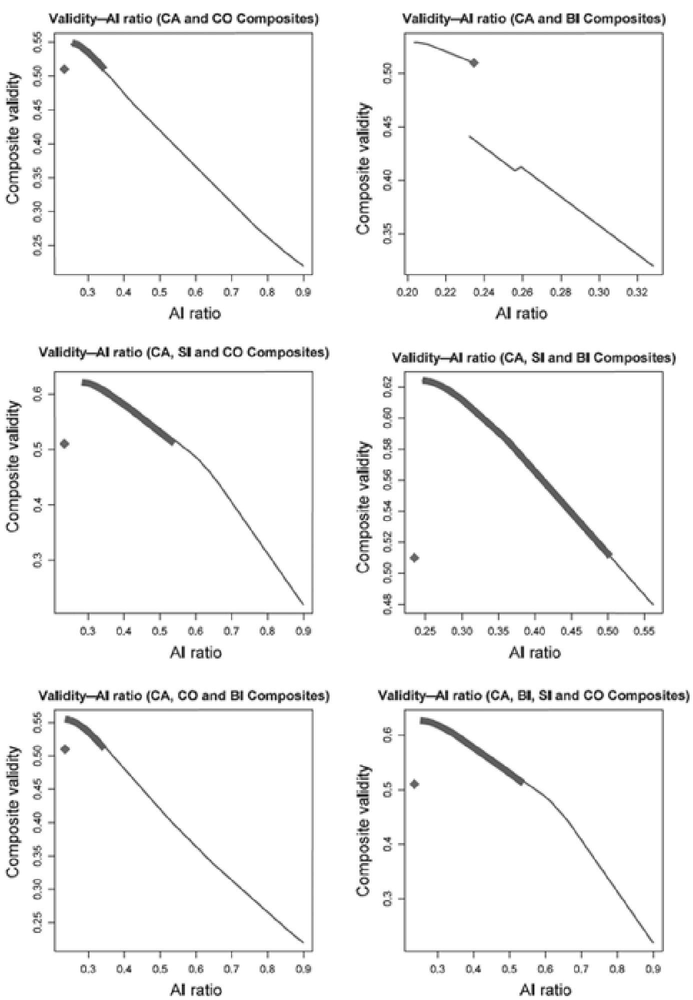

Figure 3. Pareto-optimal validity–adverse impact ratio trade-off curves for a selection with selection rate of .10 using different composites of cognitive ability and one or more non-cognitive predictors.

Table 4. AI potential of different composites at selected validity levels

Composite	Weights				Selection ratio								
CA and CO	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*
CA only	1.00	/	.00	/	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.71	/	.29	/	.55	.26	108.7 (+0)	.55	.33	107.0 (+0)	.55	.51	104.3 (+0)
99% validity	.63	/	.37	/	.54	.28	118.4 (+0)	.54	.35	114.8 (+1)	.54	.54	108.8 (+1)
97.5% validity	.58	/	.41	/	.54	.30	126.0 (+0)	.54	.37	120.8 (+1)	.54	.56	112.2 (+1)
95% validity	.54	/	.46	/	.52	.32	136.6 (+0)	.52	.40	129.0 (+1)	.52	.58	116.8 (+2)
90% validity	.47	/	.53	/	.49	.37	155.4 (+0)	.49	.45	143.3 (+2)	.49	.62	124.4 (+3)
80% validity	.37	/	.63	/	.44	.46	191.7 (+1)	.44	.54	169.9 (+3)	.44	.69	137.7 (+5)

Composite	Weights				Selection ratio								
CA and BI	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*
CA only	1.00	/	/	.00	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.75	/	/	.25	.53	.20	87.1 (−1)	.53	.27	89.3 (+0)	.53	.46	93.3 (−1)
99% validity	.87	/	/	.13	.52	.22	92.2 (−1)	.52	.29	93.6 (+0)	.52	.47	96.0 (−1)
97.5% validity	.95			.05	.52	.23	96.8 (−1)	.52	.30	97.4 (+0)	.52	.48	98.4 (−1)
95% validity	1.00	/	/	.00	.50	.24	102.7 (+0)	.50	.31	102.3 (+0)	.50	.50	101.3 (+0)
90% validity	.31	/	/	.69	.48	.24	102.2 (+0)	.48	.31	101.9 (+0)	.48	.50	101.1 (+0)
80% validity	.26	/	/	.74	.42	.25	104.9 (+0)	.42	.32	104.0 (+0)	.42	.50	102.4 (+0)

Composite	Weights				Selection ratio								
CA, SI and CO	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*
CA only	1.00	.00	.00	/	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.48	.38	.15	/	.62	.28	118.9 (+0)	.62	.36	115.2 (+1)	.62	.54	109.0 (+1)
99% validity	.38	.44	.18	/	.62	.32	133.4 (+0)	.62	.39	126.6 (+1)	.62	.57	115.4 (+2)
97.5% validity	.33	.48	.19	/	.61	.34	144.2 (+0)	.61	.42	134.8 (+1)	.61	.60	119.9 (+3)
95% validity	.27	.52	.21	/	.59	.38	158.6 (+0)	.59	.46	145.7 (+2)	.59	.63	125.6 (+4)
90% validity	.18	.58	.24	/	.56	.44	183.7 (+1)	.56	.52	164.1 (+3)	.56	.68	134.9 (+5)
80% validity	.03	.68	.29	/	.50	.56	230.4 (+2)	.50	.63	197.0 (+4)	.50	.76	150.3 (+7)

Composite	Weights				Selection ratio								
CA, SI and BI	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*	*v*	*a*	*f (n/i)*
CA only	1.00	.00	/	.00	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.43	.42	/	.16	.62	.25	104.5 (+0)	.62	.32	103.7 (+0)	.62	.50	102.2 (+0)
99% validity	.38	.52	/	.10	.62	.28	119.0 (+0)	.62	.36	115.3 (+1)	.62	.54	109.1 (+1)
97.5% validity	.36	.58	/	.07	.61	.31	129.9 (+0)	.61	.38	123.8 (+1)	.61	.56	113.9 (+2)
95% validity	.32	.65	/	.02	.59	.34	144.5 (+0)	.59	.42	135.1 (+1)	.59	.60	120.0 (+3)
90% validity	.23	.76	/	.00	.56	.41	169.9 (+1)	.56	.48	154.1 (+2)	.56	.65	129.9 (+4)
80% validity	.05	.95	/	.00	.50	.53	215.9 (+1)	.50	.60	186.9 (+3)	.50	.74	145.7 (+7)

Composite	Weights				Selection ratio								
CA, CO and BI	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f* (*n*/*i*)	*v*	*a*	*f* (*n*/*i*)	*v*	*a*	*f* (*n*/*i*)
CA only	1.00	/	.00	.00	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.64	/	.24	.12	.56	.23	99.3 (+0)	.56	.31	99.4 (+0)	.56	.49	99.7 (−1)
99% validity	.63	/	.33	.04	.55	.26	111.8 (+0)	.55	.34	109.5 (+0)	.55	.52	105.7 (+0)
97.5% validity	.61		.39	.00	.54	.29	121.1 (+0)	.54	.36	117.0 (+1)	.54	.54	110.0 (+1)
95% validity	.55	/	.45	.00	.53	.31	132.5 (+0)	.53	.39	125.9 (+1)	.53	.57	115.1 (+2)
90% validity	.48	/	.52	.00	.50	.36	151.9 (+0)	.50	.44	140.7 (+2)	.50	.61	123.0 (+3)
80% validity	.37	/	.63	.00	.44	.46	188.7 (+1)	.44	.53	167.7 (+3)	.44	.69	136.6 (+5)

Composite	Weights				Selection ratio								
CA, SI, CO, BI	CA	SI	CO	BI	.10			.20			.50		
					v	*a*	*f* (*n*/*i*)	*v*	*a*	*f* (*n*/*i*)	*v*	*a*	*f* (*n*/*i*)
CA only	1.00	.00	.00	.00	.51	.23	100.0 (2)	.51	.31	100.0 (4)	.51	.49	100.0 (16)
100% validity	.42	.37	.11	.11	.63	.26	109.0 (+0)	.63	.33	107.3 (+0)	.63	.51	104.4 (+0)
99% validity	.38	.42	.15	.05	.62	.30	125.6 (+0)	.62	.37	120.5 (+1)	.62	.55	112.0 (+1)
97.5% validity	.35	.46	.18	.01	.61	.33	137.8 (+0)	.61	.40	130.0 (+1)	.61	.58	117.3 (+2)
95% validity	.29	.50	.21	.00	.60	.37	153.6 (+0)	.60	.44	141.9 (+2)	.60	.62	123.7 (+3)
90% validity	.19	.57	.24	.00	.57	.43	179.5 (+1)	.57	.51	161.1 (+2)	.57	.67	134.4 (+5)
80% validity	.04	.67	.29	.00	.50	.55	226.8 (+1)	.50	.62	194.5 (+4)	.50	.76	149.1 (+7)

Table 4 summarizes the results of the analyses for the different predictor combinations over a representative set of overall selection rates and highlights for each combination the AI and PMR potential of balanced Pareto-optimal composites that show a 100%, 99%, 97.5%, 95%, 90% and 80% validity potential of the regression-based maximum for this combination. The corresponding expected number of selected minority applicants for a selection with a total of 500 applicants (CA only composite) and the expected gain in the number of selected minority applicants when using the other composites instead of the CA only composite is indicated as well. The reported results confirm that adding the BI predictor hardly improves the potential of the resulting composite. Also, composite formation from CA, SI and CO seems to offer the best potential, although the difference with composite formation from CA and only SI is not really substantial for comparable levels of validity. In addition it is again verified that regression-based composites seriously underestimate the potential of composite formation to balance the concerns of validity and AI. Throughout the studied predictor sets, more balanced Pareto-optimal composites offer substantial AI reduction potential at validity levels that are nearly equal to the maximum validity of the corresponding regression composite. This is also reflected by the values of both the indices *f* and *n/i*, which converge on the same pattern of findings. Consider, for example, the 95% validity composite of CA, SI and CO as compared with the CA only predictor when both are used with a selection rate of .20. In that case, four minority hires are expected when using only CA (i.e., *n* equals 4 in the .20 selection rate condition of the first row of the third subtable of Table 4) whereas two additional minority hires are expected when using the composite (cf. *i*=+2 in the fifth row of the same subtable). When looking at the corresponding *f* values, essentially the same story unfolds as these values (i.e., 100 and 145.7 for the CA only and the 95% validity composite, respectively) indicate that the expected hiring rate of the minority applicants when using the composite is 45.7% higher than that predicted under CA only selection. This is not an isolated finding because both ways of expressing the expected gain in minority hires necessarily converge on the same conclusion for large applicant samples and any difference between the two indices reflects in essence only round-off error which will be larger for smaller applicant groups with lower minority representation and lower overall selection rates.

Above it is noted that the results concerning composites from CA with either two or all three non-cognitive predictors require some caution because they are based on estimates of the corrected correlation between non-cognitive predictors that are each derived from a single study. Substituting other, quite different estimates has little effect on the previous reported results, however. This is illustrated in Figure 4 which represents the Pareto-optimal validity–AIR trade-off curve for CA, SI and CO as well as for CA, BI, SI, and CO composites as obtained when using (a) the in Table 1 reported values for the non-cognitive predictor intercorrelations (cf. the solid line in the plots), (b) substantially lower values for these intercorrelations (i.e., subtracting .15 from the Table 1 values; cf. the upper dashed line in the plots) and (c) substantially higher values for the non-cognitive predictor correlations (i.e., adding .15 to the Table 1 values; cf. the lower dashed line in the plots). The figure clearly shows that the AI and validity potential of the composites remains essentially the same for quite varying levels of the non-cognitive predictor correlations. The main difference is that lower level values for these correlations lead to a slight upward shift of the trade-off curve whereas the reverse is the case for higher level values of the correlations; a finding which merely illustrates the well-known fact that a set of predictors with lower intercorrelations enables better prediction than a similar set of more highly correlated but otherwise equally valid predictors. So, the present evidence favors the conclusion that the results reported in Table 4 will remain valid when the presently chosen values for the non-cognitive predictor correlations are replaced by more stable, meta-analytically derived estimates once such estimates become available.

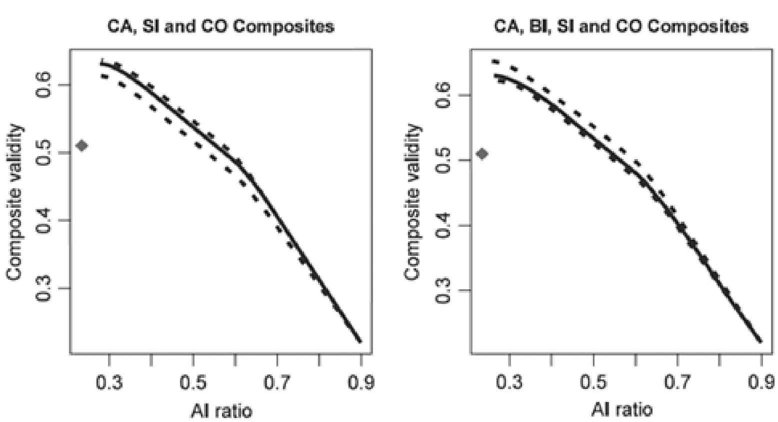

Figure 4. Sensitivity of Pareto-optimal validity–adverse impact ratio trade-off curves for varying values of the intercorrelation between non-cognitive predictors.

4. Discussion

Previous research on the AI and validity potential of composite formation focused on the merits of regression-based or *ad hoc* composites. We indicated that the regression-based composite represents only one element from the total set of Pareto-optimal composites and that it, therefore, provides only limited information on the validity and AI reduction potential of composite formation. When both validity and AI are of concern other Pareto-optimal composites may provide a better benchmark to decide on the merits of the composite formation. We also outlined a method to determine the set of Pareto-optimal composites and applied the method to a representative collection of selection predictors. The application shows that the assessment of the AI and the validity potential of composite formation can differ substantially from that arrived at when considering only regression-based composites. Thus, when studying the same predictors, but considering only regression-based composite formation, Potosky *et al.* (2005) conclude that AI was not greatly reduced by adding non-cognitive predictors, whereas we found the opposite for Pareto-optimal composites that more evenly balance the concerns of validity and AI.

Our results have the important practical implication that composite formation may often offer an alternative way out of the selection quality–AI quandary provided that one is willing to give up validity maximization as the only concern. We note that these findings regarding composite formation do not diminish the importance of other approaches to the validity–AI trade-off question, such as banding, adaptation of the presentation format or the content of tests, modification of test taker's attitudes and so on. Instead, composite formation may, in combination with these alternative approaches provide a still better answer to the vexing selection quality–AI problem than is possible by each of the alternatives alone. Also, although using predictor composites will lead to a more costly selection process as compared with using only the CA predictor, this higher cost may be more than balanced by the higher average quality of the selected applicants because the composites are usually more valid than the CA predictor (cf. the composite validities reported in Table 4). Thus, despite the higher selection costs the utility of the selection may increase.

For reasons of clarity, we described the new method to assess the merits of composite formation within a rather narrow context, considering only single-stage selections in which the applicants belong to either the majority or the minority population. In addition, we focused only on the validity and the AI reduction potential of the composite formation. Although previous research maintains the same scope, the scope will have to be broadened (a) when the candidate population is a mixture of majority and several minority groups (Blacks, Asians, Hispanics, etc.), (b) in case of multiple hurdle selection (cf. De Corte, Lievens, & Sackett, 2006) or (c) when other features than validity and AI, such as cost considerations, are relevant in comparing the potential of single predictors vs composite predictors.

Obviously, using a more balanced Pareto-optimal composite will lead to a somewhat different selected workforce than the one obtained when using the regression-based composite. The difference between the selected workforces will usually be small, however, as the more balanced composites typically have a high to very high correlation with their corresponding regression-weighted composite (e.g., Bobko *et al.*, 2007). Thus, for the presently considered predictors we found that all of the 90% or higher maximum validity composites showed at least a .85 correlation with their corresponding regression-based composite. The correlations between the composites (either the regression-based or another high maximum validity Pareto-optimal composite) and the CA predictor are typically smaller, however. So, although predictor composites are usually more valid than the CA predictor alone, using these composites, instead of the CA only predictor may sometimes lead to rather substantial changes in the make-up of the selected workforce. In particular, because composites de-emphasize the importance of CA, they may favor the hiring of candidates that show less ability to learn and/or to solve organizational problems. Selection practitioners should attend to such potential and possibly unwanted consequences. This is best achieved by being very clear about the actual type of job performance criterion that one intends to maximize by means of the selection and using predictor validity and effect size data that are relevant for the intended criterion as a basis to explore the potential of composite formation.

As a final concern it might be observed that our proposal does not dictate a particular composite or subset of composites as the generic benchmark to assess the merits of composite formation as is done in the current practice where only the regression-based composite is used. Instead, we proposed that all Pareto-optimal composites (and only those composites) are of potential interest when one is willing to value both validity and AIR and that additional considerations must be brought into play to choose any particular element or subset from the set of Pareto-optimal composites. These additional considerations reflect a value statement on the particular kind of balance between AI and validity one is aiming at. Our method does not require such value statements but rather helps to recognize the implications of these eventual concerns in terms of the still achievable trade-offs between validity and AI. The fact that the application of the method does not depend on a particular choice of the benchmark composites should therefore not be perceived as a concern but rather as strength.

In summary, researchers who study the potential of composite formation should realize that composites at best result in one of many possible Pareto-optimal trade-offs between selection quality and AI. Invariably choosing the trade-off that corresponds to the regression-based composite may often be at variance with the aim pursued by the composite formation. Finally, provided that one is willing to consider a balanced treatment of the validity–AI concerns, the results indicate that composite formation may offer a substantially better validity and AIR potential than hitherto proposed.

References

Bobko, P., Roth, P.L. and Buster, M.A. (2007) The Usefulness of Unit Weights in Creating Composite Scores. Organizational Research Methods, 10, 689–709.

Bobko, P., Roth, P.L. and Potosky, D. (1999) Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternative Predictors, and Job Performance. Personnel Psychology, 52, 561–589.

Cortina, J.M., Goldstein, N.B., Payne, S.C., Davison, H.K. and Gilliland, S.W. (2000) The Incremental Validity of Interview Scores Over and Above Cognitive Ability and Conscientiousness Scores. Personnel Psychology, 53, 325–351.

Dalessio, A. and Silverhart, T. (1994) Combining Biodata Test and Interview Information: Predicting decisions and performance criteria. Personnel Psychology, 47, 303–315.

Dana, J. and Dawes, R.M. (2004) The Superiority of Simple Alternatives to Regression for Social Science Prediction. Journal of Educational and Behavioral Statistics, 29, 317–331.

De Corte, W., Lievens, F. and Sackett, P. (2006) Predicting Adverse Impact and Multistage Mean Criterion Performance in Selection. Journal of Applied Psychology, 91, 523–537.

De Corte, W., Lievens, F. and Sackett, P. (2007) Combining Predictors to Achieve Optimal Trade-Offs Between Selection Quality and Adverse Impact. Journal of Applied Psychology, 92, 1380–1393.

Doverspike, D., Winter, J.L., Healy, M.C. and Barrett, G.V. (1996) Simulations as a Method of Illustrating the Impact of Differential Weights on Personnel Selection Outcomes. Human Performance, 9, 259–273.

Hattrup, K. and Rock, J. (2002) A Comparison of Predictor Based and Criterion Based Methods for Weighing Predictors to Reduce Adverse Impact. Applied HRM Research, 7, 22–38.

Hoffman, C.C. and Thornton, G.C. III (1997) Examining Selection Utility where Competing Predictors Differ in Adverse Impact. Personnel Psychology, 50, 455–470.

Hough, L.M., Oswald, F.L. and Ployhart, R.E. (2001) Determination, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, evidence and lessons learned. International Journal of Selection and Assessment, 9, 152–194.

Keeney, R.L. and Raiffa, H. (1993) Decisions with Multiple Objectives. Cambridge: Cambridge University Press.

Pareto, W. (1906) Manuali di Economia Politica. Milano, Italy: Societa Editrice Libraria. Translated into English by Schwier, A. S., 1971. Manual of Political Economy. New York: Macmillan.

Ployhart, R.E., Weekley, J.A., Holtz, B.C. and Kemp, C. (2003) Web-Based and Paper-and-Pencil Testing of Applicants in a Proctored Setting: Are personality, biodata, and situational judgment tests comparable? Personnel Psychology, 56, 733–752.

Potosky, D., Bobko, P. and 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 Assessment and Selection, 13, 304–315.

Sackett, P.R. and Ellingson, J.E. (1997) The Effects of Forming Multi-Predictor Composites on Group Differences and Adverse Impact. Personnel Psychology, 50, 707–721.

Sackett, P.R., Schmitt, N., Ellingson, J.E. and Kabin, M.B. (2001) High-Stakes Testing in Employment, Credentialing and Higher Education. American Psychologist, 56, 302–318.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L. and Jennings, D. (1997) AI and Predictive Efficiency of Various Predictor Combinations. Journal of Applied Psychology, 82, 719–730.