Online Supplement Tables

Table S1

Meta-Analytic Validity Estimates for Contextual Performance

Construct	Reference	Mean observed r_{xy}	r_{yy} used by meta-analysis	Validity corrected for r_{yy}	*U*	RR corrected validity	Basis for RR correction	Is RR correction credible?	Current estimate of operational validity (ρ) with best available correction
Cognitive Ability	Gonzalez-Mulé et al. (2014)	0.13	0.53	0.18	0.63	0.28	Borrowed U from F. L. Schmidt et al. (2008), as the standard deviation ratio of the GMA of applicants to incumbents was not available in any of the primary studies.	No: (a) lack of information; (b) applies correction to concurrent studies	0.18
Conscientiousness†	Chiaburu et al. (2011)	0.14	0.89	0.15	0.92	0.16	Borrowed U from F. L. Schmidt et al. (2008)	No: (a) lack of information; (b) applies correction to concurrent studies	0.15
Agreeableness†	Chiaburu et al. (2011)	0.11	0.89	0.12	0.91	0.13	Borrowed U from F. L. Schmidt et al. (2008)	No: (a) lack of information; (b) applies correction to concurrent studies	0.12
Structured Interview	Wingate et al. (2025)	0.21	Sample-specific interrater reliability used	0.28	0.74	0.3	Adopted Huffcutt et al.'s (2014) estimated ratio of .74 (all studies but Lopez, 1966)	No: (a) u borrowed from Huffcutt et al. (2014) assumed distribution with not empirical basis; (b) lack of information (only one study)	0.28
Integrity Test	Van Iddekinge et al. (2012)	0.11	0.56	0.14	NA	NA	No correction: general lack of enough range restriction information	Rightly makes no correction	0.14

Note. The current meta-analytic validity estimation followed the approach of Sackett et al. (2022; see Figure 1 in page 2047 & Table 2 in page 2050); † As noted by Chiaburu et al. (2011, p. 1144), "Because of the low amount of range restriction (high u_x ratios), the results were practically the same (difference of less than .01 in nearly all cases)," and this negligible difference had no impact on our results or conclusions.

Table S2

Meta-Analytic Validity Estimates (Biodata–Contextual Performance)

Construct	Reference	Status	*n*	Mean observed r_{xy}	r_{yy}	Validity corrected for r_{yy}	Sample	Is RR correction required?	Current estimate of operational validity (ρ) with best available correction	
Bio (rational)	Allworth & Hesketh (1999)	published	325	0.08	0.94	0.08	Concurrent, incumbent	No: concurrent sample	0.08	
Bio (rational)	Fetzer (2004)	unpublished	153	0.085	NA	NA	Concurrent, incumbent	No: concurrent sample	0.09	
Bio (rational)	Solomonson (1999)	unpublished	236	0.35	NA	NA	Concurrent, incumbent	No: concurrent sample	0.35	
Bio (rational)	Hilliard (2000)	unpublished	12	-0.136	NA	NA	Predictive, Applicants	No: predictive but not enough information	-0.14	0.23
Bio (rational)	Schmitt et al. (2009)	published	593	0.16	0.80	0.18	Concurrent, College Student	No: concurrent sample	0.18	
Bio (rational)	Sinha et al. (2011)	published	836	0.29	0.85	0.31	Concurrent, College Student	No: concurrent sample	0.31	
Bio (empirical)	Harold et al. (2006)	published	425	0.25	0.93	0.26	Concurrent, incumbent	No: concurrent sample	0.26	
Bio (empirical)	Harold et al. (2006)	published	410	0.22	0.93	0.23	Predictive, Applicants	No: predictive but not enough information	0.23	0.25

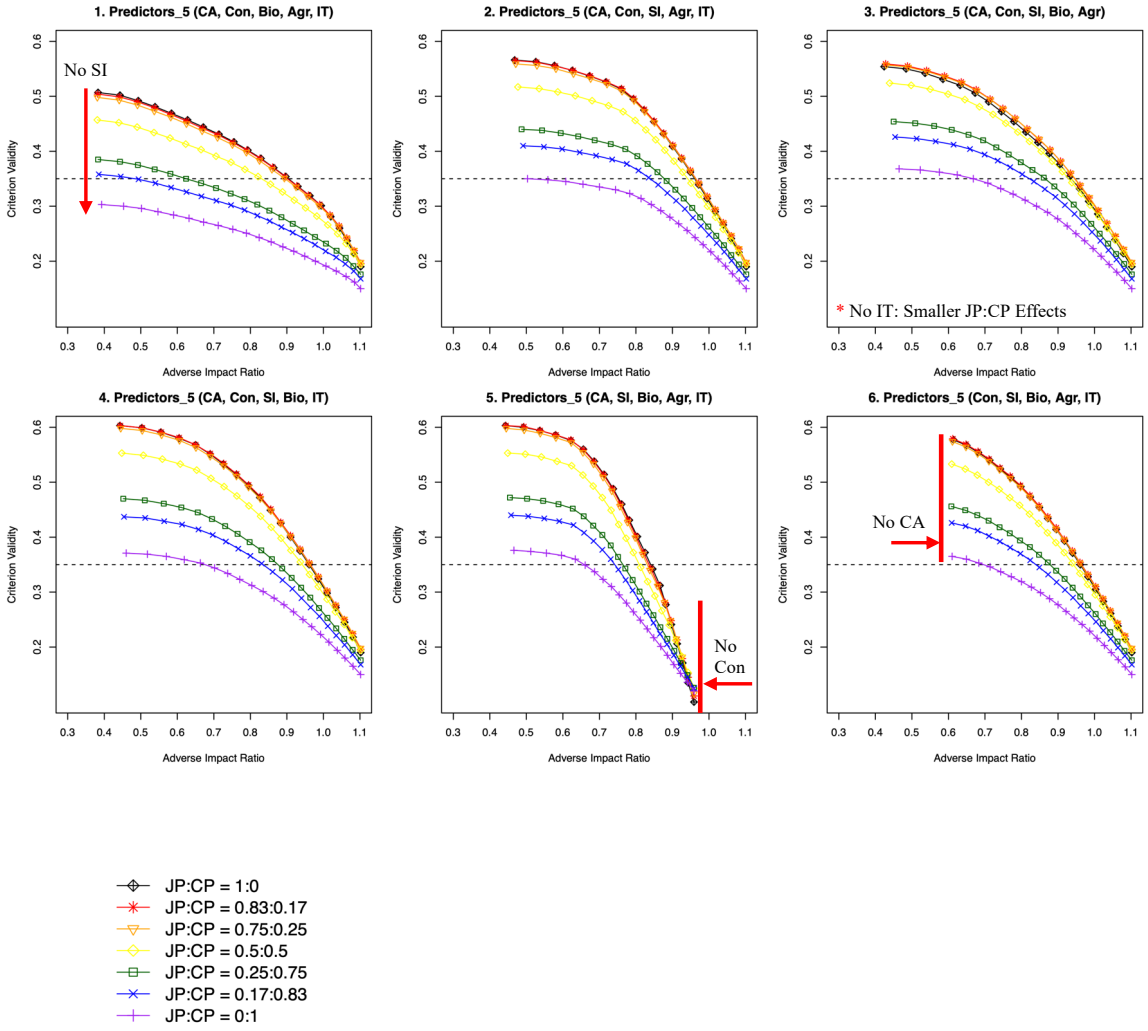
Note. To calculate the meta-analytic validity estimates between biodata and contextual performance, we relied on the meta-analytic study by Speer, Tenbrink, Wegmeyer, Sendra, Shihadeh, and Kaur (2022). We examined 109 of the 111 studies included in Speer et al. (2022), which were accessible in full text. From these, we identified studies that used contextual performance or organizational citizenship behavior (OCB) as criteria. Eight studies met this criterion—six used rationally keyed biodata and two used empirically keyed biodata. Regardless of type, the meta-analytic validities are fairly similar. However, to maintain consistency with Sackett et al.'s (2022) treatment of biodata (empirical), we opted to use 0.25 ($k = 2$, $n = 835$) for further analysis.

Online Supplement A.

Pareto-Optimal Tradeoff Curves Based upon Updated Correlation Matrix with All Combinations of Predictor Subsets (56 predictor subset combinations), and Summary of Patterns Based upon Types of Predictors Included

Figure A1

All Possible 5-Predictor Scenarios



Note. CA = cognitive ability test; Con = conscientiousness; Bio = biodata; SI = structured interview; IT = integrity test; Agr = agreeableness; JP = job performance; CP = contextual performance.

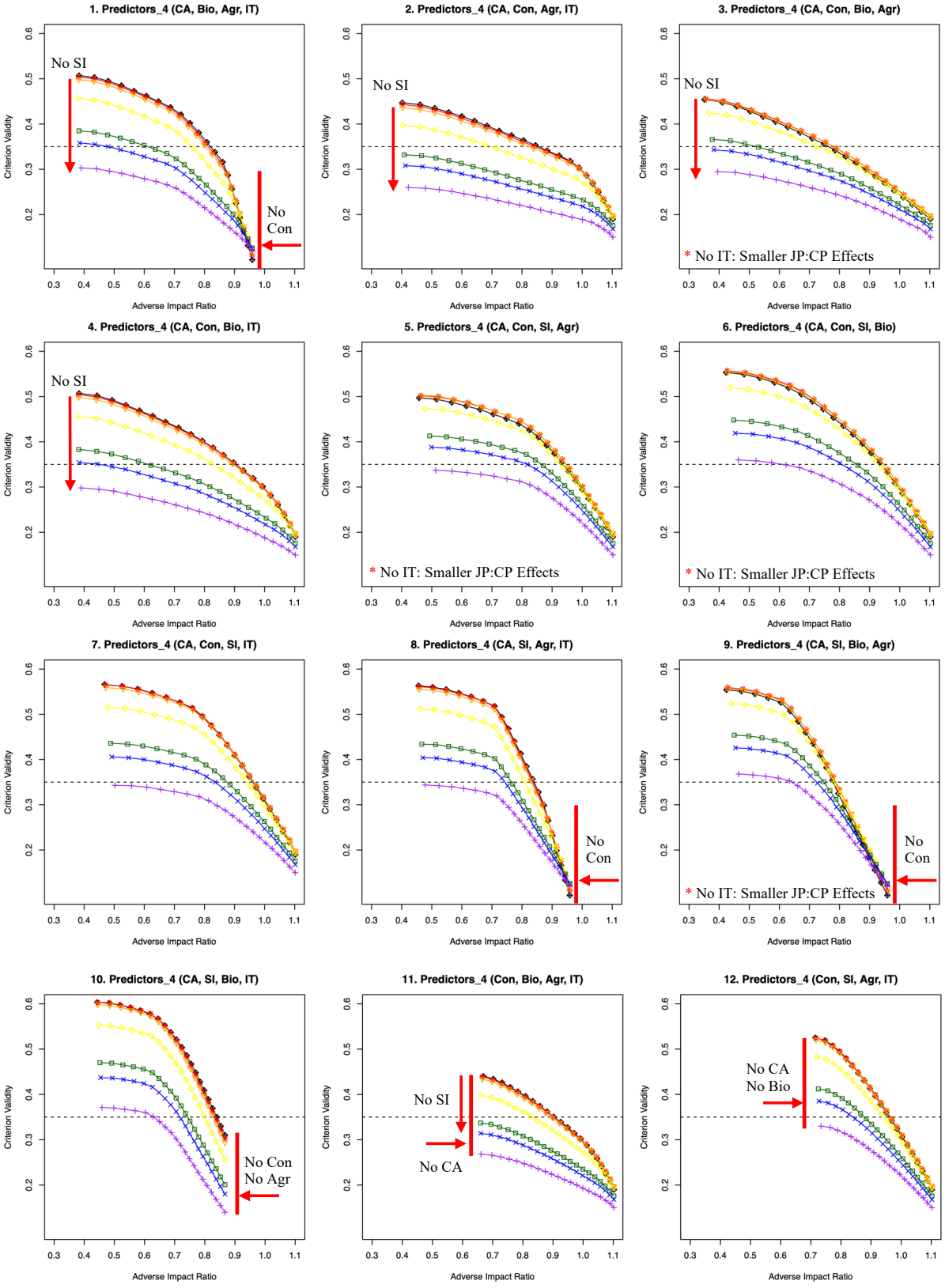
These are the patterns we observe from all the 5-predictor scenarios in Figure A1 above:

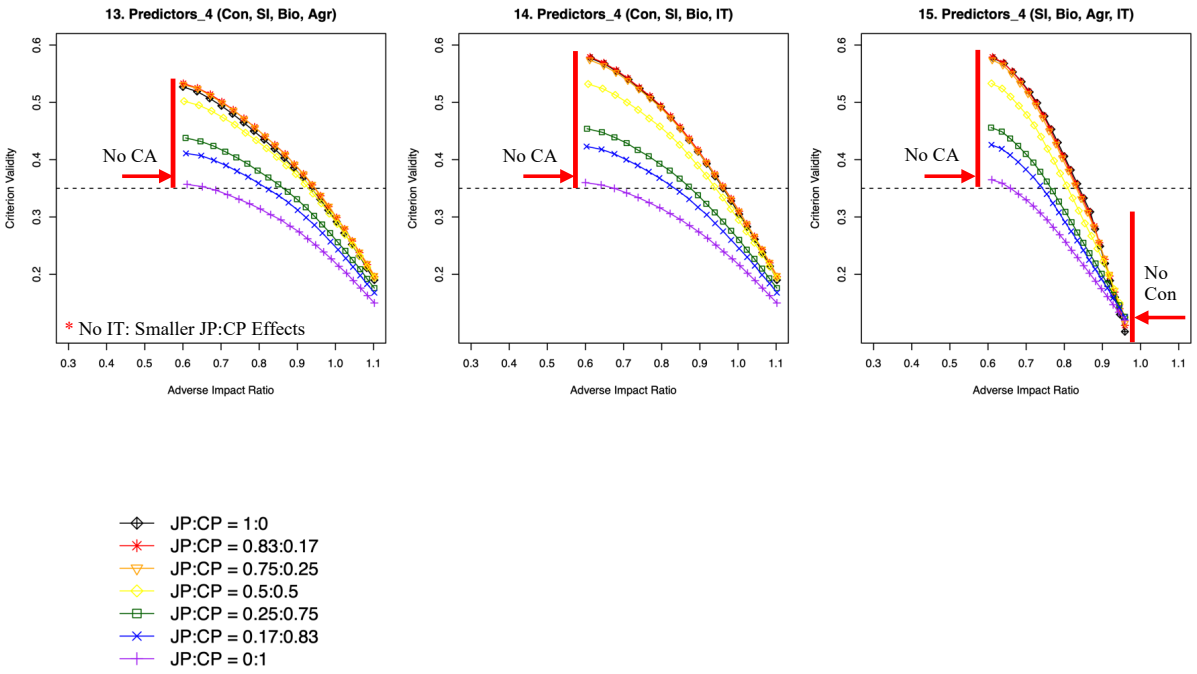
Four patterns:
Panel 1 (dropping Structured Interview): Causes the left hand side of all the curves to shift downward slightly (slightly lower validity). This happens because assigning weight to the structured interview generally improves criterion validity in most of the PO solutions (see Figure 4b).
Panel 3 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects). This happens because the weight assigned to Integrity varies across JP:CP criterion weighting conditions more than for any other predictor (e.g., see Figure 4b), likely because Integrity tests show the largest discrepancy between their JP criterion validity (r = .31) and their CP criterion validity (r = .14; discrepancy = .31 - .14 = .17). When a test predicts JP and CP very differently, its weight is more likely to change across JP:CP criterion weighting conditions.
Panel 5 (dropping Conscientiousness): Causes right hand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup $d \leq 0$. (Without Conscientiousness, the AI ratio never gets above 1.0)
Panel 6 (dropping Cognitive test): Causes left hand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6)

Two non-patterns:
Panel 2 (dropping Biodata): Has almost no effect, because the structured interview is an excellent substitute for biodata (structured interview has the same validity as biodata, and slightly lower subgroup d, so biodata already receives little weight when the structured interview is already in the model)
Panel 4 (dropping Agreeableness): Has almost no effect, because Conscientiousness is an excellent substitute for Agreeableness (Conscientiousness has slightly higher validity and slightly lower subgroup d than Agreeableness, so Agreeableness already receives little weight when Conscientiousness is already in the model)

Figure A2

All Possible 4-Predictor Scenarios





Note. CA = cognitive ability test; Con = conscientiousness; Bio = biodata; SI = structured interview; IT = integrity test; Agr = agreeableness; JP = job performance; CP = contextual performance.

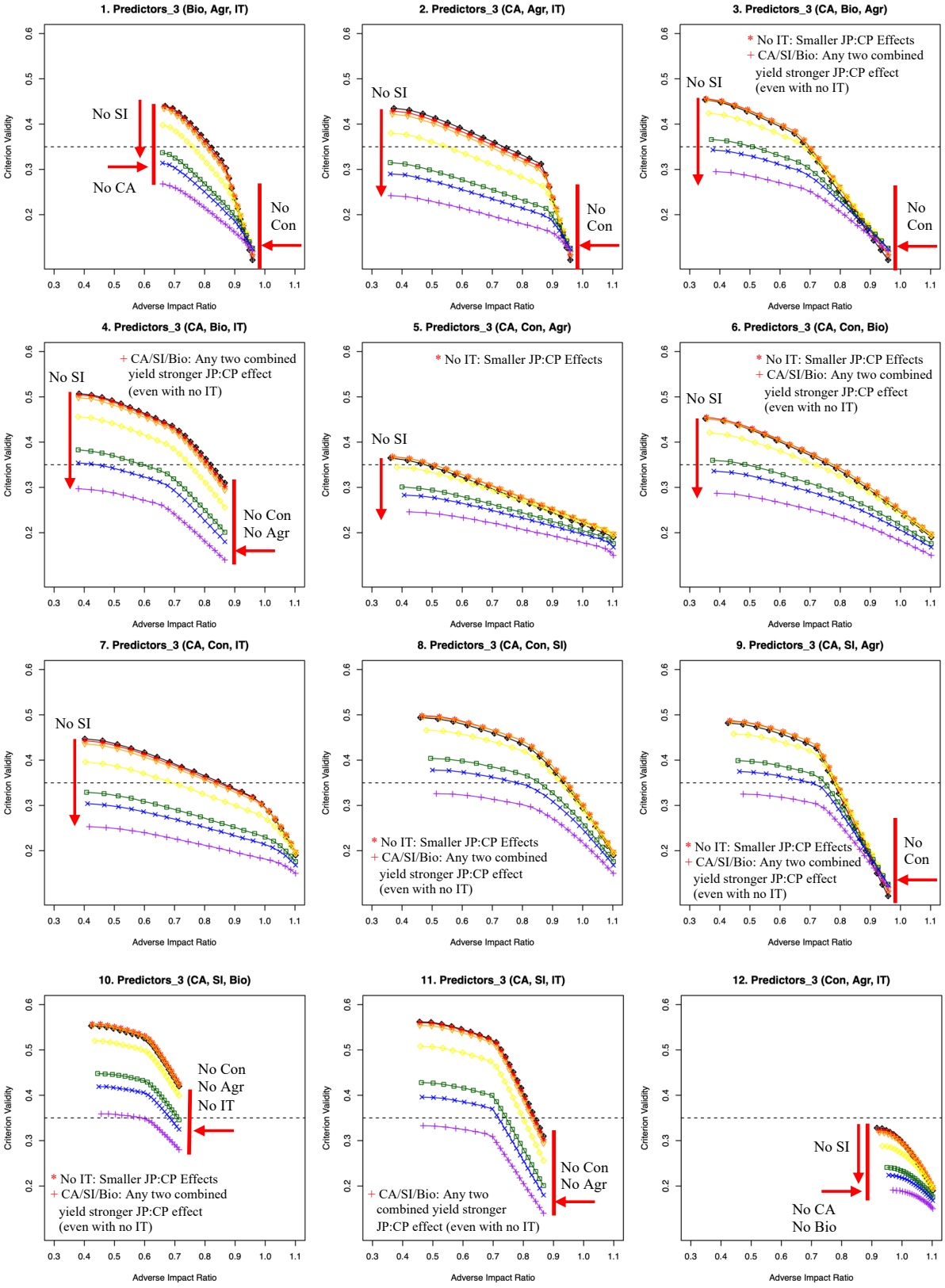
These are the patterns we observe from all the 4-predictor scenarios in Figure A2 above:

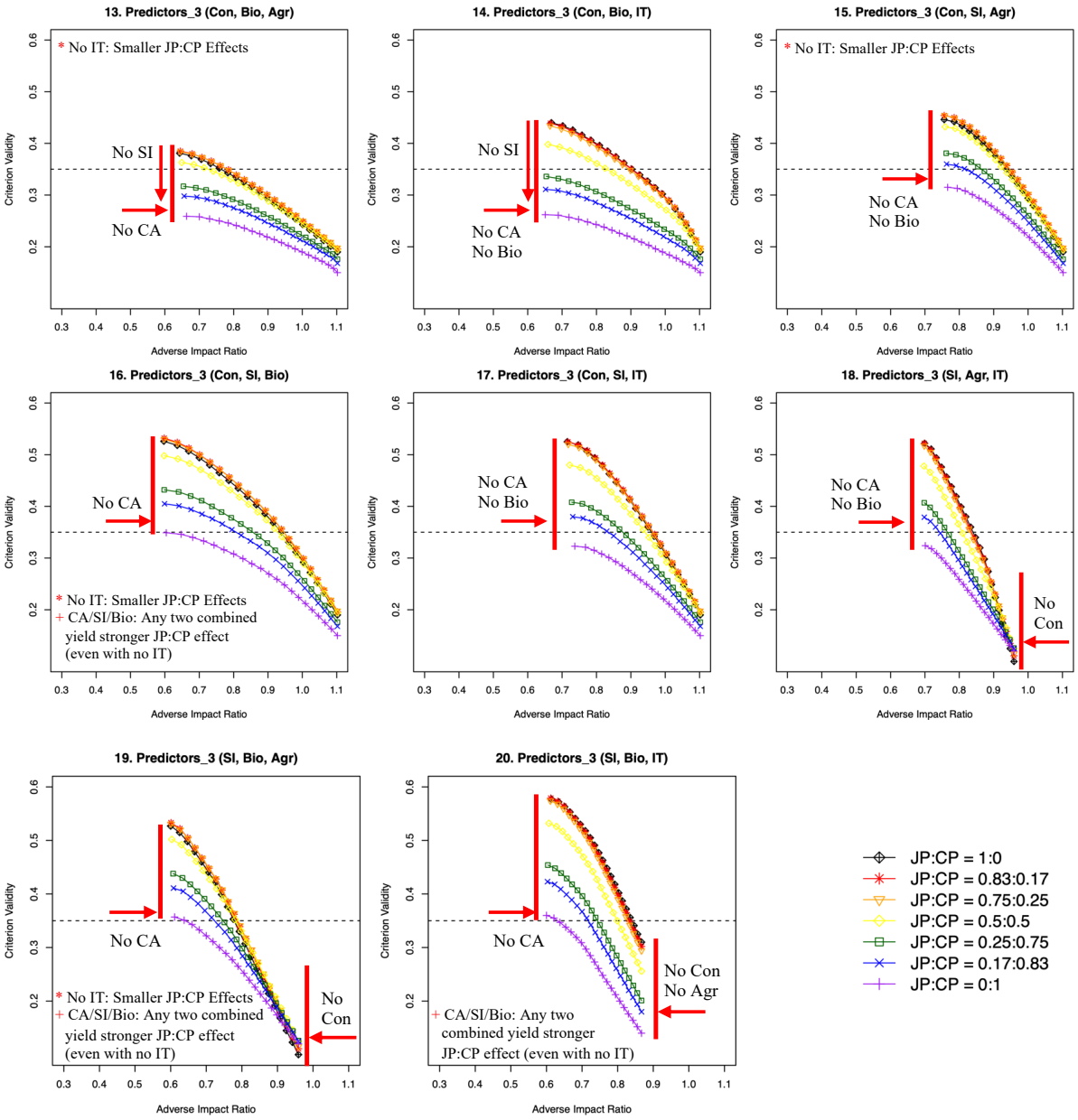
Four patterns:
Panels 1, 2, 3, 4, and 11 (dropping Structured Interview): Causes the left hand side of all the curves to shift downward slightly (slightly lower validity)
Panels 3, 5, 6, 9, and 13 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects)
Panels 1, 8, 9, 10, and 15 (dropping Conscientiousness): Causes right hand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d $\leq$ 0. (Without Conscientiousness, the AI ratio never gets above 1.0)
Panels 11, 12, 13, 14, and 15 (dropping Cognitive test): Causes left hand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6)

Two additional patterns:
Panel 10 (dropping Agreeableness and Conscientiousness): Causes right hand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Conscientiousness and also without Agreeableness, the AI ratio never gets above ~.85).
Panel 12 (dropping Biodata and Cognitive test): Causes left hand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without cognitive test and also without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test

Figure A3

All Possible 3-Predictor Scenarios





Note. CA = cognitive ability test; Con = conscientiousness; Bio = biodata; SI = structured interview; IT = integrity test; Agr = agreeableness; JP = job performance; CP = contextual performance.

These are the patterns we observe from all the 3-predictor scenarios in Figure A3 above:

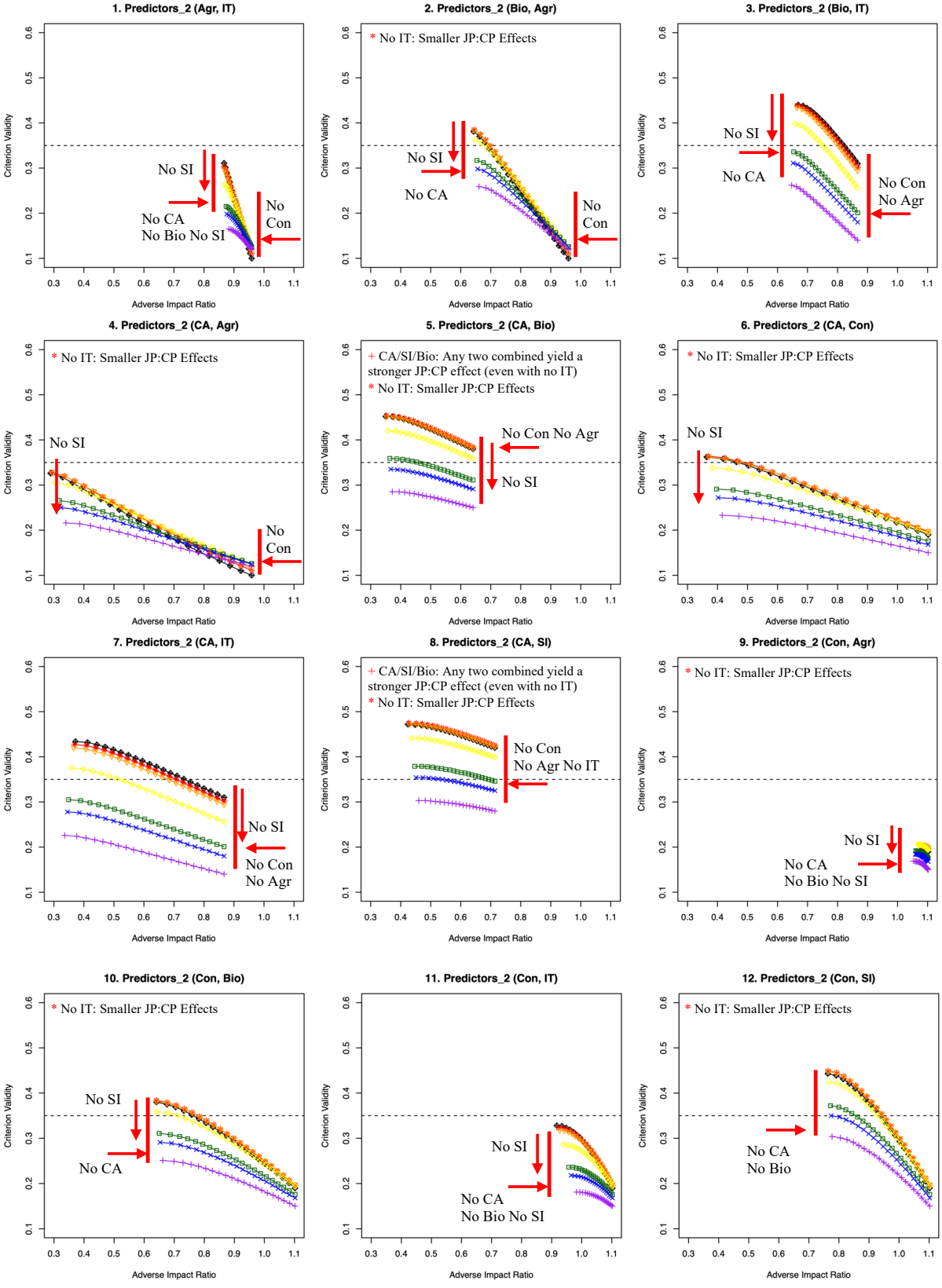
Four patterns:
Panels 1, 2, 3, 4, 5, 6, 7, 12, 13, and 14 (dropping Structured Interview): Causes the lefthand side of all the curves to shift downward slightly (slightly lower validity)

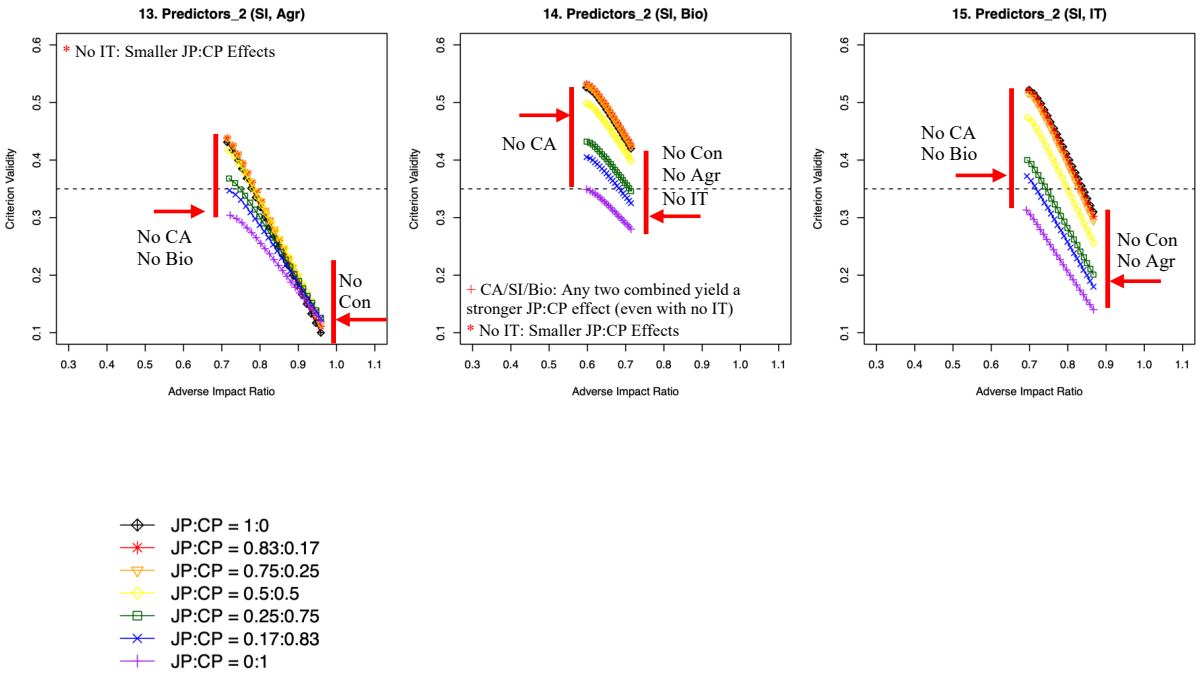
Panels 3, 5, 6, 8, 9, 10, 13, 15, 16, and 19 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects)
Panels 1, 2, 3, 4, 9, 10, 11, 18, 19, and 20 (dropping Conscientiousness): Causes righthand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d <= 0. (Without Conscientiousness, the AI ratio never gets above 1.0)
Panels 1, 12, 13, 14, 15, 16, 17, 18, 19, and 20 (dropping Cognitive test): Causes lefthand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6)

Two additional patterns:
Panels 4, 10, 11, and 20 (dropping Agreeableness and Conscientiousness): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Conscientiousness and also without Agreeableness, the AI ratio never gets above ~.85).
Panels 12, 15, 17, and 18 (dropping Biodata and Cognitive test): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without cognitive test and also without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test.

Figure A4

All Possible 2-Predictor Scenarios





Note. CA = cognitive ability test; Con = conscientiousness; Bio = biodata; SI = structured interview; IT = integrity test; Agr = agreeableness; JP = job performance; CP = contextual performance.

These are the patterns we observe from all the 2-predictor scenarios in Figure A4 above:

Four patterns:
Panels 1, 2, 3, 4, 5, 6, 7, 9, 10, and 11(dropping Structured Interview): Causes the lefthand side of all the curves to shift downward slightly (slightly lower validity)
Panels 2, 4, 5, 6, 8, 9, 10, 12, 13, and 14 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects)
Panels 1, 2, 3, 4, 5, 7, 8, 13, 14, and 15 (dropping Conscientiousness): Causes righthand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d <= 0. (Without Conscientiousness, the AI ratio never gets above 1.0)
Panels 1, 2, 3, 9, 10, 11, 12, 13, 14, and 15 (dropping Cognitive test): Causes lefthand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6)

Four additional patterns:
Panels 3, 7, and 15 (dropping Agreeableness and Conscientiousness): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Conscientiousness and also without Agreeableness, the AI ratio never gets above ~.85).
Panels 5, 8, and 14 (dropping Agreeableness and Conscientiousness and Integrity): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without Conscientiousness and also without Agreeableness and Integrity test, the AI ratio never gets above ~.7).
Panels 12, 13, and 15 (dropping Biodata and Cognitive test): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without cognitive test and also without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test

Panels 1, 9, and 11 (dropping Biodata and Cognitive test and Structured Interview): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .9 (without cognitive test and also without biodata and structured Interview, the AI ratio never gets below .85)—biodata has the second largest subgroup d after cognitive test and structured interview has the third largest subgroup d.

Table A1

Summary of Patterns

5-Predictor (Figure A1)	4-Predictor (Figure A2)
Four patterns: Panel 1 (dropping Structured Interview): Causes the left hand side of all the curves to shift downward slightly (slightly lower validity) Panel 3 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects) Panel 5 (dropping Conscientiousness): Causes right hand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d ≤ 0. (Without Conscientiousness, the AI ratio never gets above 1.0) Panel 6 (dropping Cognitive test): Causes left hand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6) **Two non-patterns**: Panel 2 (dropping Biodata): Has almost no effect, because structured interview is an excellent substitute for biodata (structured interview has the same validity as biodata, and slightly lower subgroup d, so biodata already receives little weight when structured interview is already in the model) Panel 4 (dropping Agreeableness): Has almost no effect, because Conscientiousness is an excellent substitute for Agreeableness (Conscientiousness has slightly higher validity and slightly lower subgroup d than Agreeableness, so Agreeableness already receives little weight when Conscientiousness is already in the model)	**Four patterns**: Panels 1, 2, 3, 4, and 11 (dropping Structured Interview): Causes the left hand side of all the curves to shift downward slightly (slightly lower validity) Panels 3, 5, 6, 9, and 13 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects) Panels 1, 8, 9, 10, and 15 (dropping Conscientiousness): Causes right hand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d ≤ 0. (Without Conscientiousness, the AI ratio never gets above 1.0) Panels 11, 12, 13, 14, and 15 (dropping Cognitive test): Causes left hand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6) **Two additional patterns**: Panel 10 (dropping Agreeableness and Conscientiousness): Causes right hand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Consciousness. and also without Agreeableness, the AI ratio never gets above ~.85). Panel 12 (dropping Biodata and Cognitive test): Causes left hand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without Cognitive test and also without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test
3-Predictor (Figure A3)	2-Predictor (Figure A4)
Four patterns: Panels 1, 2, 3, 4, 5, 6, 7, 12, 13, and 14 (dropping Structured Interview): Causes the lefthand side of all the curves to shift downward slightly (slightly lower validity) Panels 3, 5, 6, 8, 9, 10, 13, 15, 16, and 19 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects) Panels 1, 2, 3, 4, 9, 10, 11, 18, 19, and 20 (dropping Conscientiousness): Causes righthand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d <= 0. (Without Conscientiousness, the AI ratio never gets above 1.0) Panels 1, 12, 13, 14, 15, 16, 17, 18, 19, and 20 (dropping Cognitive test): Causes lefthand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6) **Two additional patterns**: Panels 4, 10, 11, and 20 (dropping Agreeableness and Conscientiousness): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Conscientiousness and also without Agreeableness, the AI ratio never gets above ~.85). Panels 12, 15, 17, and 18 (dropping Biodata and Cognitive test): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without cognitive test and also	**Four patterns:** Panels 1, 2, 3, 4, 5, 6, 7, 9, 10, and 11 (dropping Structured Interview): Causes the lefthand side of all the curves to shift downward slightly (slightly lower validity) Panels 2, 4, 5, 6, 8, 9, 10, 12, 13, and 14 (dropping Integrity): Causes the different colored tradeoff curves to be closer together (smaller JP:CP criterion weighting effects) Panels 1, 2, 3, 4, 5, 7, 8, 13, 14, and 15 (dropping Conscientiousness): Causes righthand side of tradeoff curves to be pushed to the left, because Conscientiousness is the only predictor that has subgroup d <= 0. (Without Conscientiousness, the AI ratio never gets above 1.0) Panels 1, 2, 3, 9, 10, 11, 12, 13, 14, and 15 (dropping Cognitive test): Causes lefthand side of tradeoff curves to be truncated around AI ratio = .6 (without cognitive test, the AI ratio never gets below .6) **Four additional patterns**: Panels 3, 7, and 15 (dropping Agreeableness and Conscientiousness): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .85 (without Conscientiousness and also without Agreeableness, the AI ratio never gets above ~.85). Panels 5, 8, and 14 (dropping Agreeableness and Conscientiousness and Integrity): Causes righthand side of tradeoff curves to be even more truncated, around AI ratio = .7

without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test	(without Conscientiousness and also without Agreeableness and Integrity test, the AI ratio never gets above ~.7). Panels 12, 13, and 15 (dropping Biodata and Cognitive test): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .7 (without cognitive test and also without biodata, the AI ratio never gets below .7)—biodata has the second largest subgroup d, after cognitive test Panels 1, 9, and 11 (dropping Biodata and Cognitive test and Structured Interview): Causes lefthand side of tradeoff curves to be even more truncated, around AI ratio = .9 (without cognitive test and also without biodata and structured interview, the AI ratio never gets below .85)—biodata has the second largest subgroup d after cognitive test, and structured interview has the third largest subgroup d.

The table above describes the overall patterns we found in the 5-predictor, 4-predictor, 3-predictor, and 2-predictor scenarios. There are four main patterns:

(Pattern #1) dropping the structured interview causes the left-hand side of all the curves to shift downward slightly;
(Pattern #2) dropping the integrity test causes the different colored tradeoff curves to move closer together, indicating smaller JP:CP criterion weighting effects;
(Pattern #3a) dropping Conscientiousness truncates the right-hand side of the tradeoff curves, (Pattern #3b) dropping both Conscientiousness and Agreeableness results in further truncation of the right-hand side of the curves, around an AI ratio of .85, and (Pattern #3c) dropping Conscientiousness and Agreeableness and the integrity test results in further truncation of the right-hand side of the curves, around an AI ratio of .7; and
(Pattern #4a) dropping the cognitive test causes the left-hand side of the tradeoff curves to be truncated around an AI ratio of .6, (Pattern #4b) dropping both cognitive test and biodata causes the left-hand side of the tradeoff curves to be even more truncated, around an AI ratio of .7, and (Pattern #4c) dropping the cognitive test and biodata and the structured interview causes the left-hand side of the tradeoff curves to be even more truncated, around an AI ratio of .9.

We also identified two non-patterns in the 5-predictor case: (1) dropping biodata has almost no effect, because the structured interview serves as an excellent substitute for biodata; and (2) dropping Agreeableness has little impact, as Conscientiousness is an effective substitute.

Quantifying Improvements in AI Ratio and Criterion Validity due to Joint Optimization of Predictor and Criterion Weights (Outer Envelope Approach) across All Predictor Subsets

Finally, we present simple metrics in order to quantify the average improvement in AI ratio (ΔAI Ratio$_{avg}$) and average improvement in criterion validity (ΔR_{avg}), due to using the outer envelope/jointly optimized predictor and criterion weights—versus using the gray tradeoff curve (criterion weighting alone) and versus using the black tradeoff curve (PO predictor weighting alone). These results are given in Table A2 below, for all 56 predictor subsets as well as the Hattrup et al. (1997) dataset. Table A2 shows three results. First, across all predictor subsets based on the updated meta-analytic correlation matrix, the average ΔAI Ratio$_{avg(env\ vs\ CWalone)}$

= .159, and $\Delta R_{avg(env\ vs\ CWalone)}$ = .061, showing that jointly optimizing criterion and predictor weights (the outer envelope approach) offers substantial advantage beyond criterion weighting alone (Hattrup et al.'s 1997 approach). This trend showing the superiority of the outer envelope approach over criterion weighting alone is also seen in Hattrup et al.'s own input data matrix, where $\Delta AI\ Ratio_{avg(env\ vs\ CWalone)}$ = .148 and $\Delta R_{avg(env\ vs\ CWalone)}$ = .097. Combining criterion weighting with PO predictor weighting (outer envelope) routinely outperforms criterion weighting alone (gray curve).

Second, across all predictor subsets based on the updated meta-analytic correlation matrix, the average $\Delta AI\ Ratio_{avg(env\ vs\ POalone)}$ = .008, and the average $\Delta R_{avg(env\ vs\ POalone)}$ = .005 (both metrics are below .01). In short, jointly optimizing criterion and predictor weights (outer envelope approach) offers little advantage beyond optimizing predictor weights alone (the black curve). That is, criterion weighting offers little benefit, after PO predictor weighting is already being used.

These lackluster results for criterion weighting in the updated meta-analytic matrix stand in stark contrast to results from Hattrup et al.'s (1997) older input data matrix, for which $\Delta AI\ Ratio_{avg(env\ vs\ POalone)}$ = .223 and $\Delta R_{avg(env\ vs\ POalone)}$ = .135 (i.e., criterion weighting substantially improves upon PO alone in Hattrup et al.'s 1997 data). This raises the question of why criterion weighting augments PO predictor weighting in Hattrup et al.'s (1997) input data, but does not do so in the updated meta-analytic data. Inspired by Hattrup et al.'s (1997) older input data, we next conducted supplemental simulations to better pinpoint data circumstances where criterion weighting is likely to produce the most benefit (see Online Appendix B).

Table A2. *Average Gain in AI Ratio and Criterion Validity due to Using Global Pareto Frontier/Outer Envelope vs. Criterion Weighting Alone and PO Predictor Weighting Alone, for all 56 Predictor Subsets and the Hattrup et al. 1997 Data*

Predictor Subset	Outer Envelope/Global Pareto Frontier minus Criterion Weighting alone (Gray Curve) [Outer Envelope vs. Criterion Weighting Alone]			Outer Envelope/Global Pareto Frontier minus Pareto-Optimal Predictor Weighting alone (Black Curve) [Outer Envelope vs. PO Predictor Weighting Alone]	
	$\Delta AI\ Ratio_{avg(env\ vs\ CWalone)}$	$\Delta R_{avg(env\ vs\ CWalone)}$		$\Delta AI\ Ratio_{avg(env\ vs\ POalone)}$	$\Delta R_{avg(env\ vs\ POalone)}$
CA, Con, SI, Bio, Agr, IT	.230	.100		.004	.003
Con, SI, Bio, Agr, IT	.150	.000		.004	.003
CA, SI, Bio, Agr, IT	.190	.100		.002	.002
CA, Con, Bio, Agr, IT	.270	.120		.002	.001
CA, Con, SI, Agr, IT	.240	.090		.004	.003
CA, Con, SI, Bio, IT	.240	.100		.004	.003
CA, Con, SI, Bio, Agr	.210	.080		.013	.007
SI, Bio, Agr, IT	.100	.000		.002	.003
Con, Bio, Agr, IT	.170	.000		.002	.001
Con, SI, Agr, IT	.110	.080		.004	.003
Con, SI, Bio, IT	.150	.000		.004	.003

Con, SI, Bio, Agr	.130	.070		.011	.008
CA, Bio, Agr, IT	.240	.120		.002	.002
CA, SI, Agr, IT	.200	.100		.002	.002
CA, SI, Bio, IT	.190	.100		.001	.001
CA, SI, Bio, Agr	.160	.080		.013	.012
CA, Con, Agr, IT	.310	.080		.002	.001
CA, Con, Bio, IT	.280	.120		.002	.001
CA, Con, Bio, Agr	.220	.070		.017	.006
CA, Con, SI, IT	.240	.100		.004	.003
CA, Con, SI, Agr	.260	.080		.015	.008
CA, Con, SI, Bio	.220	.080		.013	.007
Bio, Agr, IT	.110	.000		.002	.003
SI, Agr, IT	.060	.110		.002	.003
SI, Bio, IT	.100	.000		.000	.000
SI, Bio, Agr	.070	.070		.012	.014
Con, Agr, IT	.080	.050		.003	.002
Con, Bio, IT	.170	.000		.002	.001
Con, Bio, Agr	.150	.050		.016	.007
Con, SI, IT	.110	.090		.004	.003
Con, SI, Agr	.090	.070		.010	.008
Con, SI, Bio	.140	.070		.011	.007
CA, Agr, IT	.330	.000		.002	.001
CA, Bio, IT	.200	.000		.000	.000
CA, Bio, Agr	.180	.070		.017	.010
CA, SI, IT	.210	.100		.001	.000
CA, SI, Agr	.200	.060		.016	.012
CA, SI, Bio	.140	.080		.010	.005
CA, Con, IT	.310	.070		.002	.001
CA, Con, Agr	.230	.050		.031	.007
CA, Con, Bio	.230	.070		.015	.005
CA, Con, SI	.230	.070		.014	.007
Agr, IT	.020	.060		.003	.009
Bio, IT	.090	.000		.000	.000
Bio, Agr	.060	.050		.017	.015
SI, IT	.070	.000		.000	.000
SI, Agr	.040	.060		.013	.017
SI, Bio	.040	.070		.005	.005
Con, IT	.070	.050		.003	.002
Con, Agr	.020	.010		NA	.010
Con, Bio	.140	.050		.015	.006
Con, SI	.090	.070		.010	.008
CA, IT	.120	NA		.000	.000
CA, Agr	.110	.040		.043	.015
CA, Bio	.060	.070		.012	.003
CA, SI	.080	.070		.021	.004
CA, Con	.220	.050		.024	.006
Average across all predictor subsets (updated meta-analytic data)	**.159**	**.061**		**.008**	**.005**
Hattrup et al.'s (1997) older input data	.148	.097		.223	.135

Online Supplement B.

Exploratory simulations to consider different input data matrices, inspired by Hattrup et al.'s (1997) older input data

The simulations below are designed to pinpoint conditions when criterion weighting might possibly be beneficial. Hattrup et al.'s (1997) older data illustrated that, despite the finding that criterion weighting is not likely to be beneficial on average (based upon all subsets of the updated meta-analytic input matrix; see Figure 3a, Figure 4a, and Appendix A), there might exist edge conditions where criterion weighting could be combined with PO predictor weighting to improve the diversity-validity outcome pairings. That is, Hattrup's old input data matrix inspired us to conduct simulations to find patterns where—although potentially uncommon—criterion weighting might be beneficial.

Simulation #1: Case with cognitive test + low-impact predictor (i.e., Conscientiousness or Integrity or Agreeableness).

We began by simulating the common 2-predictor scenario in which a cognitive test is combined with a low-impact predictor, such as Conscientiousness, integrity, or Agreeableness. Beginning with the updated meta-analytic estimates in Table 2, we extracted the following Simulation #1 Correlation Matrix. Note how the job performance criterion validity of the low-impact predictor is manipulated to be .1 (similar to Agreeableness), .2 (similar to Conscientiousness), or .3 (similar to integrity test); and the subgroup d of the low-impact predictor is manipulated to be -.1 (similar to Conscientiousness), 0 (similar to Agreeableness), or .1 (similar to integrity test).

Table B1
Correlation Matrix: Inspired by Cognitive Test + Conscientiousness or Integrity or Agreeableness

Variable	1	2	3	4
1. Cognitive ability	1			
2. Conscientiousness or Integrity or Agreeableness	0	1		
3. Job (task) performance	.3	.1,.2,.3	1	
4. Contextual performance	.2	.15	.5	1
Black-White *d*	.8	-.1,0,.1	-	-

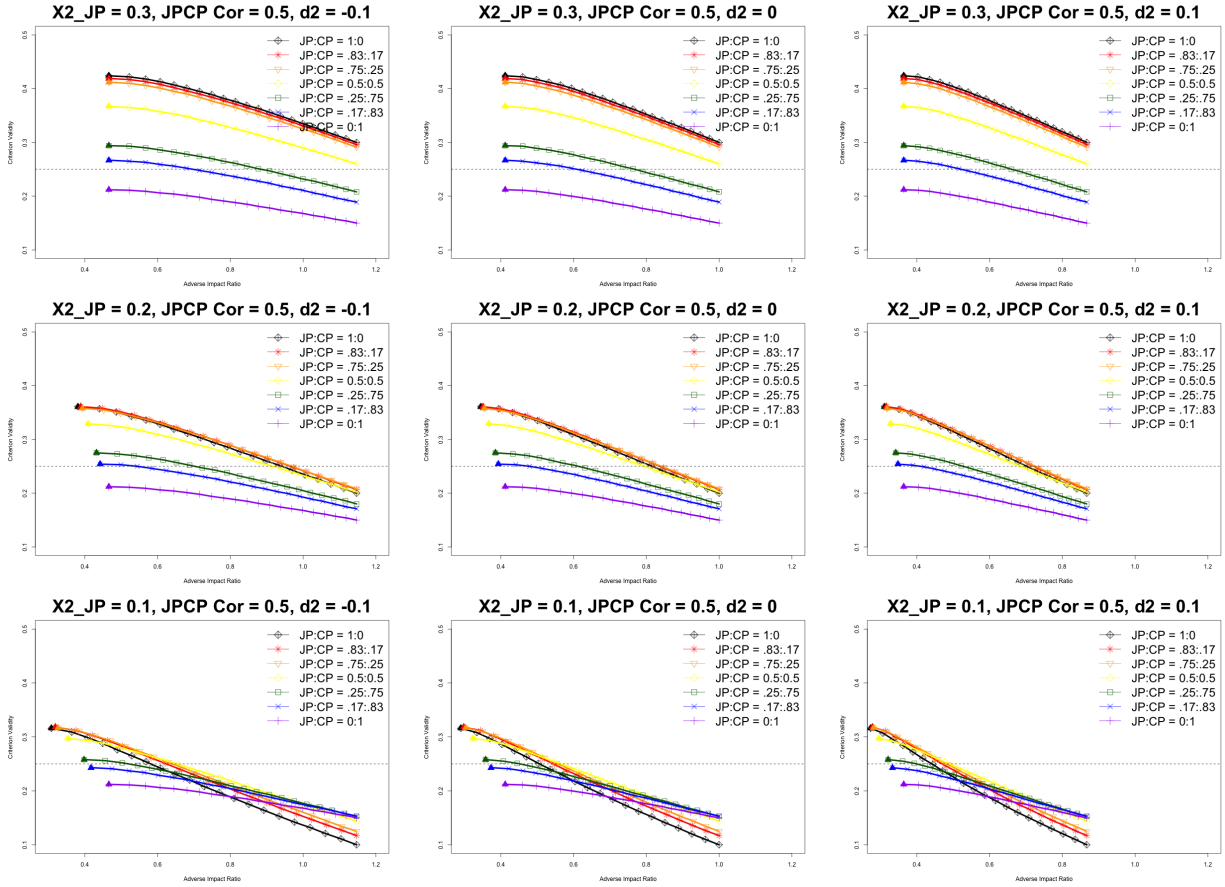
Next, we created CW+PO tradeoff curve plots based upon the Simulation #1 correlation matrix (Figure B1).

In Figure B1 below:

- The job performance criterion validity of X2 (i.e., $r_{X2,JP}$) decreases as you move down a column.
- Subgroup d of X2 increases as you move to the right.

Figure B1

Case with Cognitive Test + low-impact predictor



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

The above simulation reveals two trends:

1. "Hattrup Condition": $r_{X2,CP} > r_{X2,JP}$ enables CW+PO to sometimes outperform CW alone and PO alone. Moving down a column (from $r_{X2,JP} = .3$ down to $r_{X2,JP} = .1$), we see that when $r_{X2,JP}$ falls < $r_{X2,CP}$ (i.e., when $r_{X2,JP} < .15$ in the current example, which occurs only on the bottom row of Figure B1), we get a pattern similar to the Hattrup's old input data matrix (i.e., assigning higher weight to contextual performance flattens the PO tradeoff curve). Under such conditions (i.e., the bottom row of Figure B1, when $r_{X2,JP}$ falls below $r_{X2,CP}$) the optimal solutions do contain contextual performance criterion weighting (i.e., the colorful PO curves sometimes outperform the black PO curve—meaning CW+PO outperforms PO alone).

2. Subgroup d of X2 has little effect. Moving across a row from left to right (from d2 = -0.1 to 0 to 0.1), the rightmost ends of the tradeoff curves become truncated. For example, when d2 = 0, the rightmost ends of the tradeoff curves are truncated at AI ratio = 1.0, and when d2 = .1, the rightmost ends of the tradeoff curves are truncated at AI ratio = ~.85. Otherwise, manipulating d2 across the realistic range shown in the Simulation #1 correlation matrix (which is based on Table 2) has little effect on the PO tradeoff curves, and we find d2 has almost no implications for criterion weighting effects.

<u>Simulation #2: Atypical case with cognitive test + low-impact predictor</u> (i.e., Conscientiousness or Integrity or Agreeableness).

Because Simulation #1 revealed that $r_{X2,CP} > r_{X2,JP}$ was a key condition that makes criterion weighting work, we now use Simulation #2 to further explore the relationship between $r_{X2,JP}$ and $r_{X2,CP}$ in driving the criterion weighting effects. We thus simulated an atypical 2-predictor scenario, in which a cognitive test is combined with a low-impact predictor (e.g., Conscientiousness, integrity, or Agreeableness), but instead of using the typical contextual performance criterion validity value for X2 (i.e., instead of using $r_{X2,CP} = .15$, which is approximately true for Conscientiousness, integrity, and Agreeableness—see Table 2—we manipulated $r_{X2,CP}$ to be .1, .2, or .3).

Table B2
Correlation Matrix: Inspired by Cognitive Test + Conscientiousness or Integrity or Agreeableness (including large estimates for $r_{X2,CP}$, similar to how Hattrup et al.'s old input data had large $r_{X2,CP}$)

Variable	1	2	3	4
1. Cognitive ability	1			
2. Conscientiousness or Integrity or Agreeableness	0	1		
3. Job (task) performance	.3	.1,.2,.3	1	
4. Contextual performance	.2	.1,.2,.3	.3,.5	1
Black-White *d*	.8	0	-	-

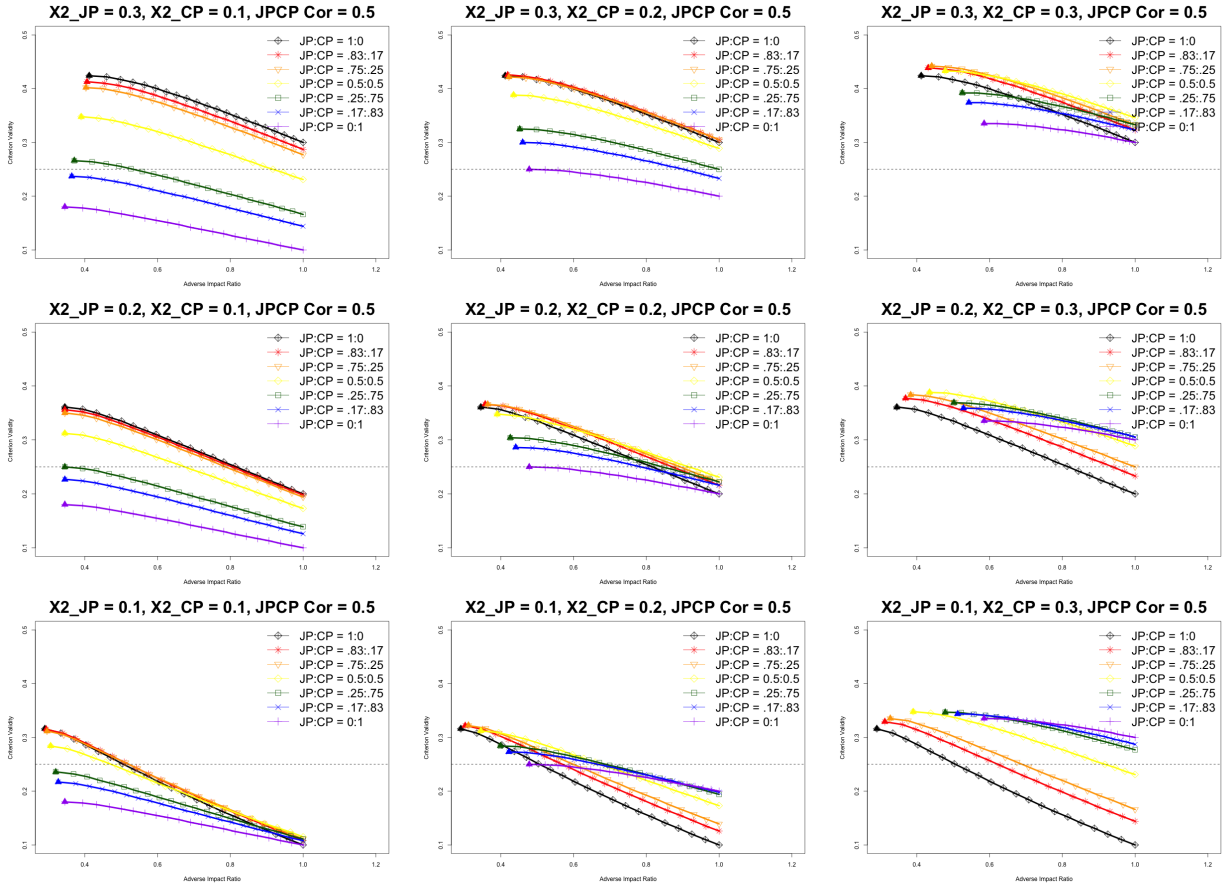
Next, we created CW+PO tradeoff curve plots based upon the Simulation #2 correlation matrix (Figure B2).

In Figure B2 below:

- Job performance criterion validity of X2 (i.e., $r_{X2,JP}$) decreases as you move down a column.
- Contextual performance criterion validity of X2 (i.e., $r_{X2,CP}$) increases as you move to the right.
- If you move both down and to the right (moving toward the lower right corner), the contextual performance criterion validity of X2 trends toward becoming larger than the job performance criterion validity of X2 (i.e., $r_{X2,CP} > r_{X2,JP}$)

Figure B2a

Atypical Case with Cognitive Test + low-impact predictor ($r_{JP,CP}$ = 0.5)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

The above simulation (Simulation #2) again reveals one major trend (echoing Simulation #1):

"Hattrup Condition": $r_{X2,CP} > r_{X2,JP}$ enables CW+PO to sometimes outperform CW alone and PO alone. Moving across a row from left to right (from $r_{X2,CP}$ = .1 up to $r_{X2,CP}$ = .3), and moving down a column (from $r_{X2,JP}$ = .3 down to $r_{X2,JP}$ = .1), we see again that when $r_{X2,CP}$ grows > $r_{X2,JP}$, we get a pattern similar to the Hattrup's old input data matrix (i.e., assigning higher weight to contextual performance flattens the PO tradeoff curve). Under such conditions (i.e., when $r_{X2,CP} > r_{X2,JP}$) the optimal solutions do contain contextual performance criterion weighting (i.e., the colorful PO curves sometimes outperform the black PO curve; CW+PO outperforms PO alone).

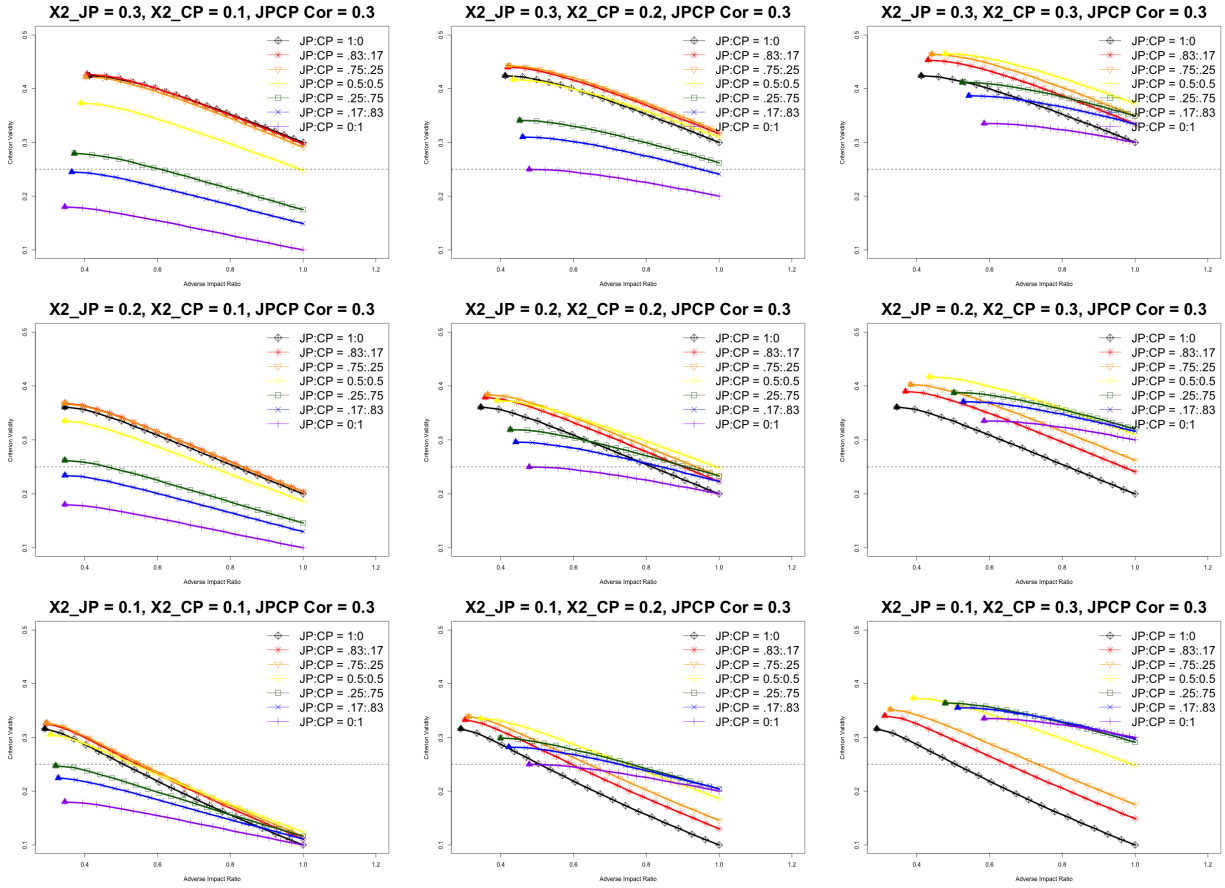
Simulation #2b. Correlation between criteria (manipulating $r_{JP,CP} = .3$)

Smaller criterion intercorrelation. As the correlation between JP and CP decreases from .5 to .3 (also reflecting an uncommon condition found in Hattrup et al.'s (1997) older data input matrix), the validity of the red and orange PO curves can rise slightly above the black PO curve, which also indicates a condition where the optimal solutions can contain contextual performance

criterion weighting (with unusually small criterion intercorrelation, CW+PO can exceed PO alone, even when $r_{X2,JP} = r_{X2,CP}$; see the diagonal (lower left to upper right) of Figure B2b below).

Figure B2b

Atypical Case with Cognitive Test + low-impact predictor ($r_{JP,CP}$=0.3)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

<u>Simulation #3: Case with structured interview + low-impact predictor</u> (i.e., Conscientiousness or Integrity or Agreeableness).

Simulations #1 and #2 above focused on using a cognitive test plus a low-impact predictor. We now additionally investigate how criterion weighting might work in the absence of a cognitive test. Another potentially common 2-predictor scenario might entail a structured interview combined with a low-impact predictor, such as Conscientiousness, integrity, or Agreeableness. Beginning with the updated meta-analytic estimates in Table 2, we extracted the following Simulation #3 Correlation Matrix.

Table B3
Correlation Matrix: Inspired by Structured Interview + Conscientiousness or Integrity or Agreeableness

Variable	1	2	3	4
1. Structured Interview	1			
2. Conscientiousness or Integrity or Agreeableness	0,.1	1		
3. Job (task) performance	.4	.1,.2,.3	1	
4. Contextual performance	.3	.15	.5	1
Black-White *d*	.25	-.1,0,.1	-	-

Next, we created CW+PO tradeoff curve plots based upon the Simulation #3 correlation matrix (Figure B3).

In Figure B3 below:
- The job performance criterion validity of X2 (i.e., $r_{X2,JP}$) decreases as you move down.
- Subgroup d of X2 increases as you move to the right.

For Structured Interview (X1) + Conscientiousness or Integrity or Agreeableness (X2), here are the trends:

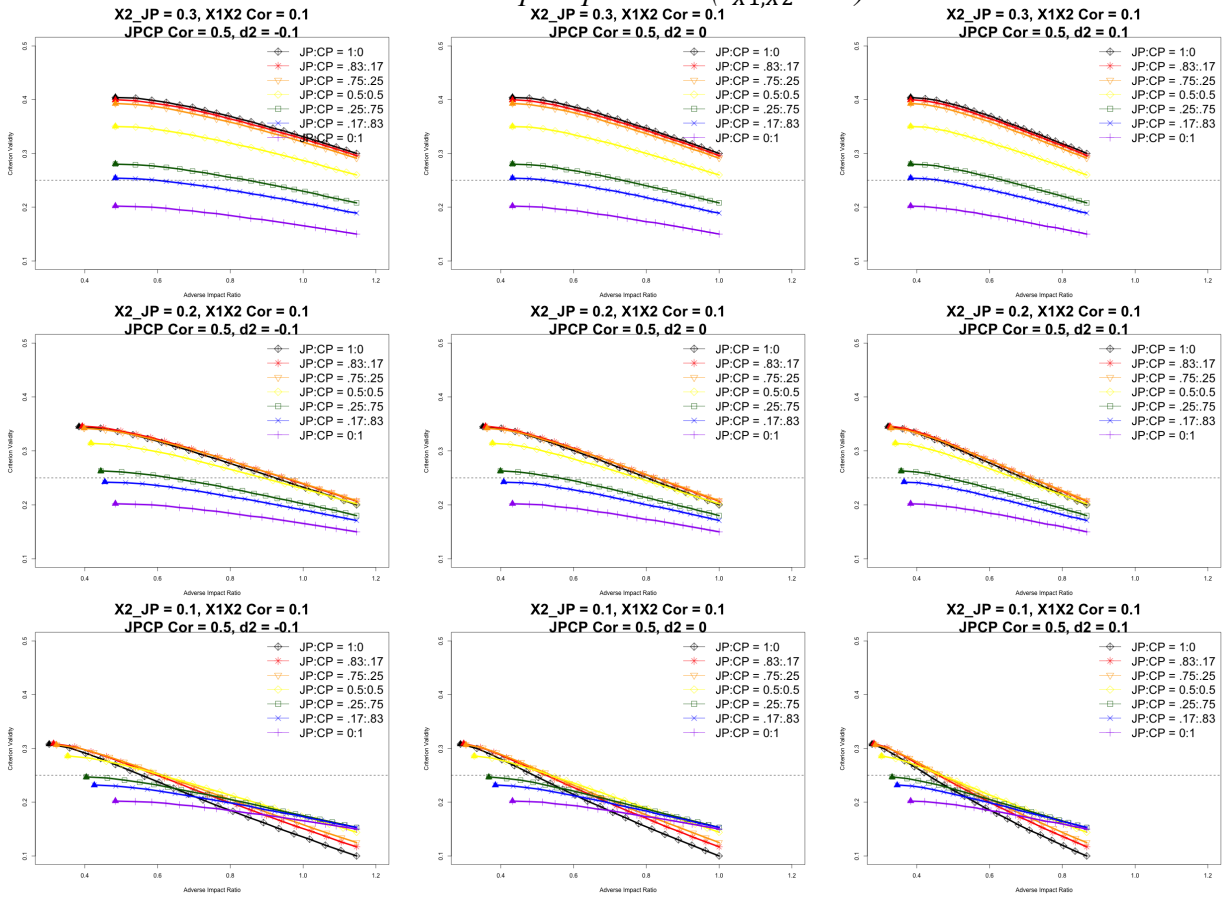
1. <u>"Hattrup Condition": $r_{X2,CP} > r_{X2,JP}$ enables CW+PO to sometimes outperform PO alone</u>. Moving down a column (from $r_{X2,JP} = .3$ down to $r_{X2,JP} = .1$), we see that when $r_{X2,JP}$ falls < $r_{X2,CP}$ (e.g., when $r_{X2,JP} < .15$ in the current example, which occurs only on the bottom row of Figure B3), we get a pattern similar to Hattrup's old input data matrix (i.e., assigning higher weight to contextual performance flattens the PO tradeoff curve). Under such conditions (i.e., the bottom row of Figure B3, when $r_{X2,JP}$ falls below $r_{X2,CP}$) the optimal solutions do contain contextual performance criterion weighting (i.e., the colorful PO curves sometimes outperform the black PO curve; CW+PO outperforms PO alone).

2. <u>Subgroup d of X2 has little effect</u>. Moving across a row from left to right (from d2 = -0.1 to 0 to 0.1), the rightmost ends of the tradeoff curves become truncated. Otherwise, d2 has little effect on the PO tradeoff curves, and d2 has almost no implications for criterion weighting effects.

3. <u>Predictor intercorrelation has little effect</u>. Under conditions inspired by Table 2, Figure B3a and Figure B3b are very similar.

Figure B3a

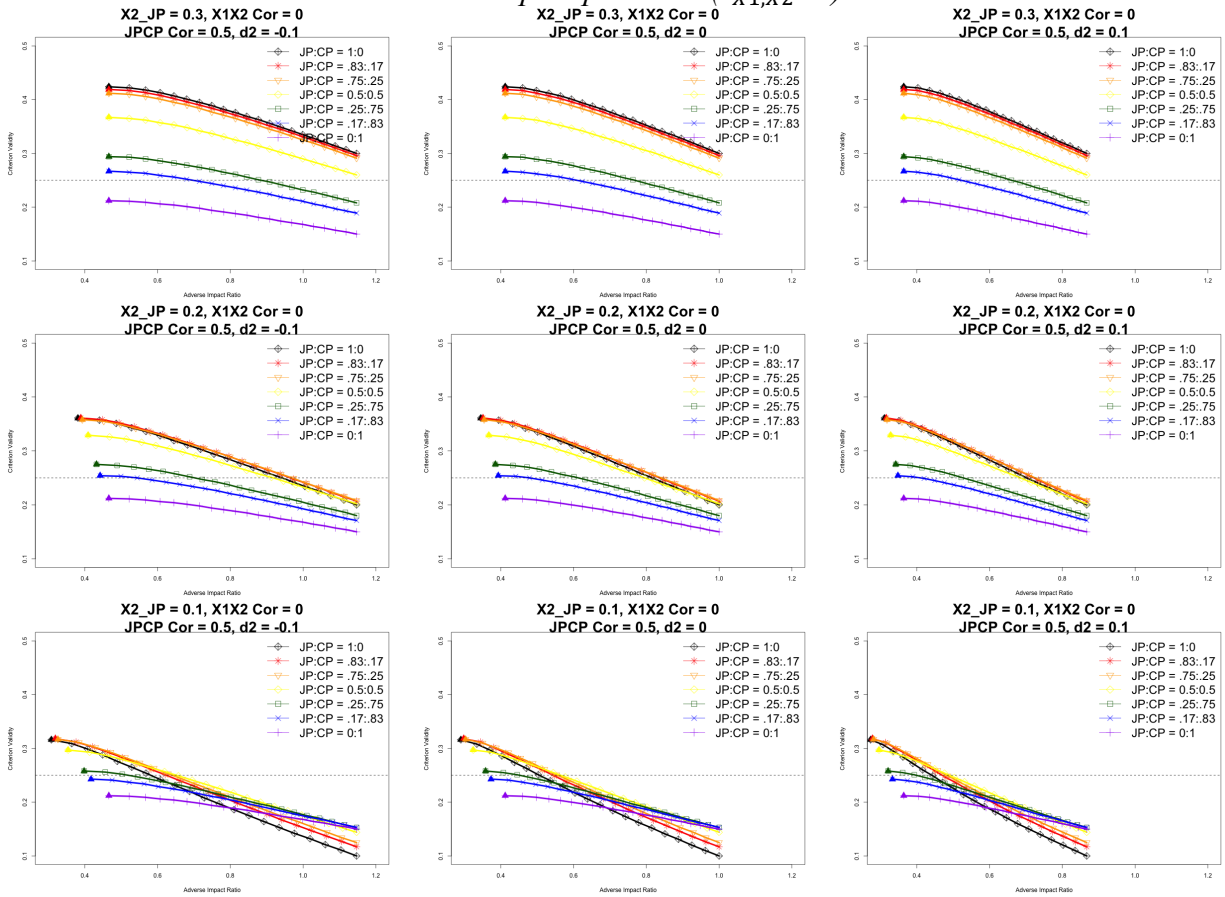
Case with Structured Interview + low-impact predictor ($r_{X1,X2}$=0.1)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

Figure B3b

Case with Structured Interview + low-impact predictor ($r_{X1,X2}$=0)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

<u>Simulation #4: Atypical case with structured interview + low-impact predictor</u> (i.e., Conscientiousness or Integrity or Agreeableness).

Because Simulation #3 revealed that $r_{X2,CP} > r_{X2,JP}$ was a key condition that makes criterion weighting work, we now use Simulation #4 to further explore the relationship between $r_{X2,JP}$ and $r_{X2,CP}$ in driving the criterion weighting effects. We thus simulated a 2-predictor scenario, in which a structured interview is combined with a low-impact predictor (e.g., Conscientiousness, integrity, or Agreeableness), but instead of using a typical contextual performance criterion validity value for X2 (i.e., instead of using $r_{X2,CP} = .15$, which is approximately true for Conscientiousness, integrity, and Agreeableness [see Table 2]; we manipulated $r_{X2,CP}$ to be .1, .2, or .3).

Table B4
Correlation Matrix: Inspired by Structured Interview + Conscientiousness or Integrity or Agreeableness (including large estimates for $r_{X2,CP}$ similar to how Hattrup et al.'s old input data had large $r_{X2,CP}$)

Variable	1	2	3	4
1. Structured Interview	1			
2. Conscientiousness or Integrity or Agreeableness	0,.1	1		
3. Job (task) performance	.4	.1,.2,.3	1	
4. Contextual performance	.3	.1,.2,.3	.3,.5	1
Black-White *d*	.25	0	-	-

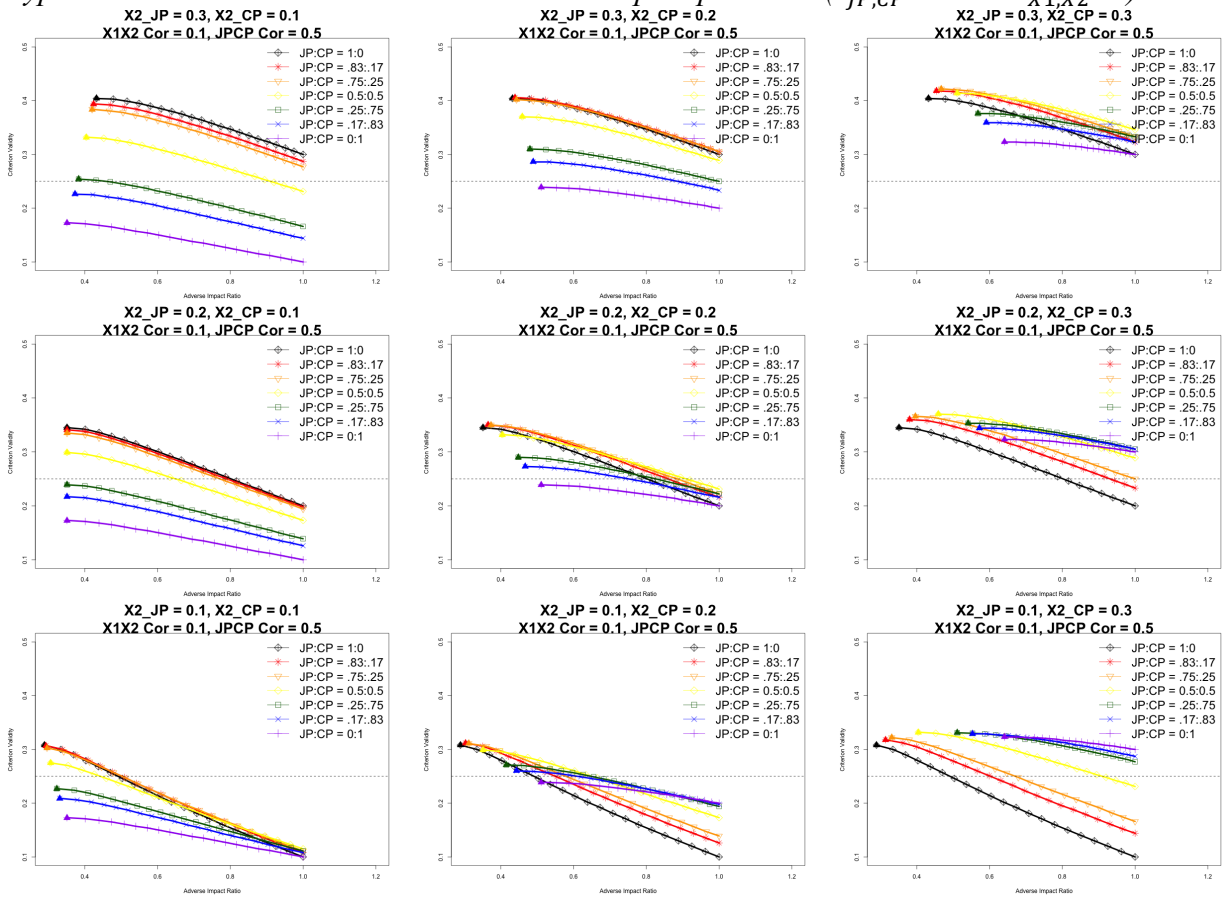
Next, we created CW+PO tradeoff curve plots based upon the Simulation #4 correlation matrix (Figure B4).

In Figure B4 below:

- Job performance criterion validity of X2 (i.e., $r_{X2,JP}$) decreases as you move down.
- Contextual performance criterion validity of X2 (i.e., $r_{X2,CP}$) increases as you move to the right.
- If you move both down and to the right (moving toward the lower right corner), contextual performance criterion validity of X2 trends toward becoming larger than job performance criterion validity of X2 (i.e., $r_{X2,CP} > r_{X2,JP}$).

Figure B4a

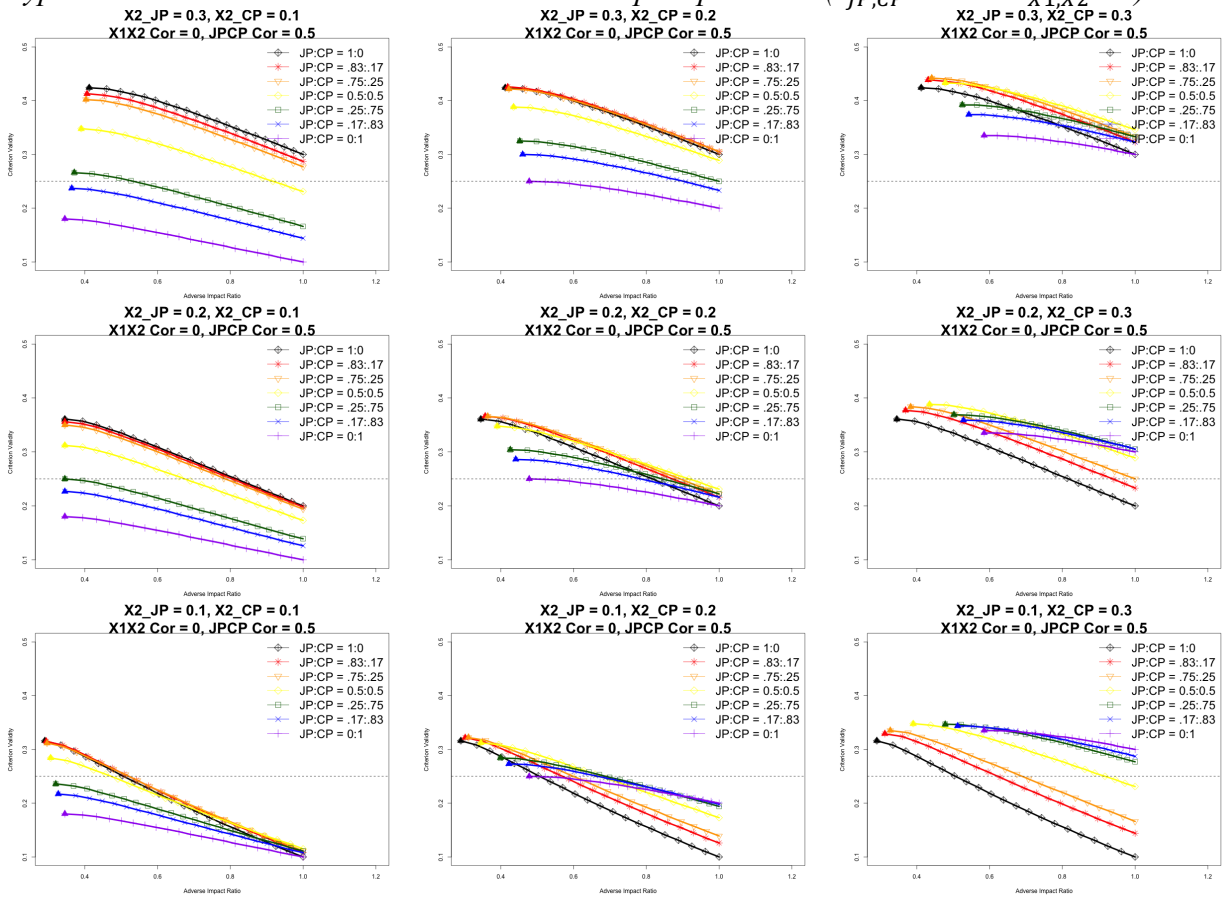
Atypical Case with Structured Interview + low-impact predictor ($r_{JP,CP}$=.5 and $r_{X1,X2}$=0)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

Figure B4b

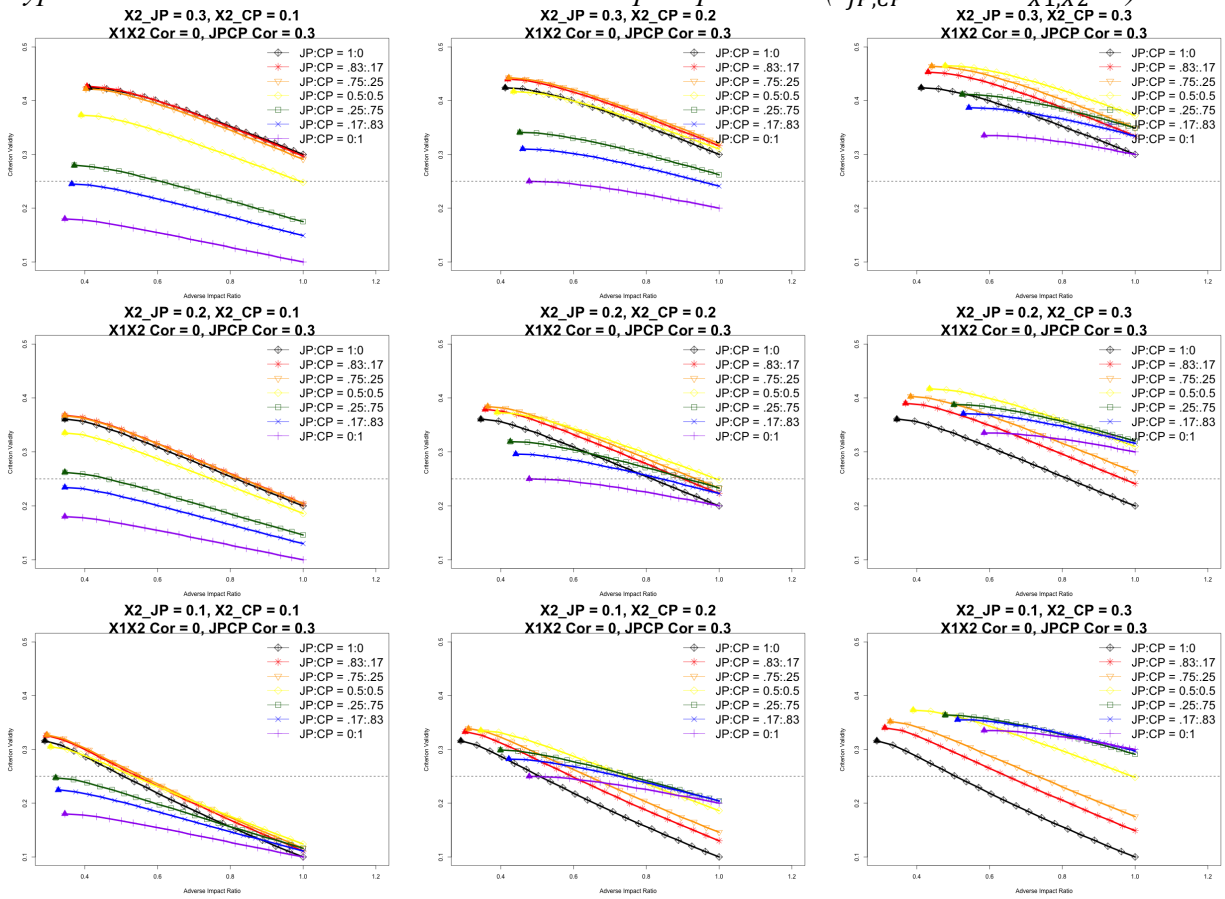
Atypical Case with Structured Interview + low-impact predictor ($r_{JP,CP}$=.5 and $r_{X1,X2}$=.1)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

Figure B4c

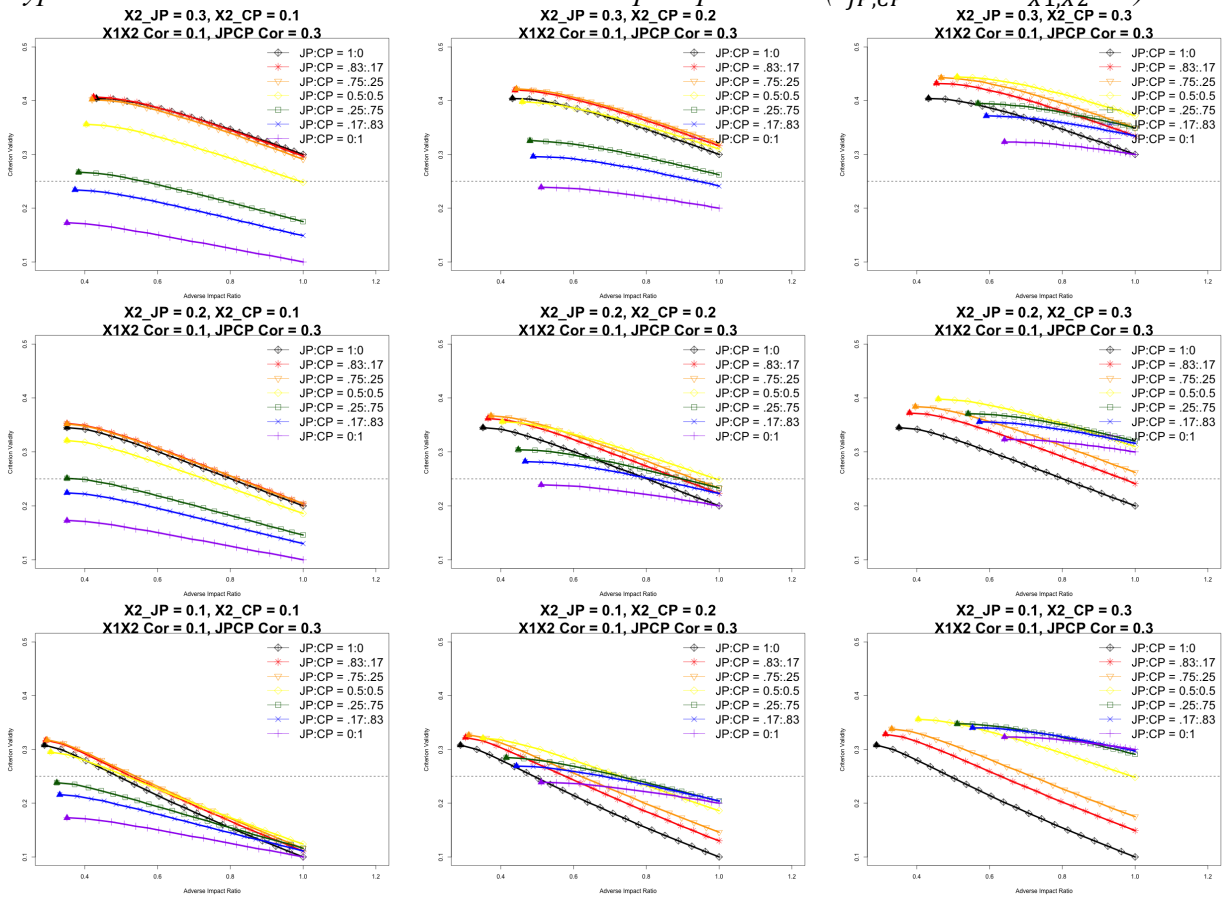
Atypical Case with Structured Interview + low-impact predictor ($r_{JP,CP}$=.3 and $r_{X1,X2}$=0)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

Figure B4d

Atypical Case with Structured Interview + low-impact predictor ($r_{JP,CP}$=.3 and $r_{X1,X2}$=.1)



Note. X2_JP = job performance criterion validity of the low-impact predictor; X2_CP = contextual performance criterion validity of the low-impact predictor; X1X2 Cor = predictor intercorrelation; JPCP Cor = correlation between criteria.

The above simulation (Simulation #4) again reveals one major trend (echoing Simulations #1, #2, and #3):

<u>"Hattrup Condition": $r_{X2,CP} > r_{X2,JP}$ enables CW+PO to sometimes outperform PO alone</u>. Moving across a row from left to right (from $r_{X2,CP}$ = .1 up to $r_{X2,CP}$ = .3), and moving down a column (from $r_{X2,JP}$ = .3 down to $r_{X2,JP}$ = .1), we see again that when $r_{X2,CP}$ grows > $r_{X2,JP}$, we get a pattern similar to the Hattrup's old input data matrix (i.e., assigning higher weight to contextual performance flattens the PO tradeoff curve). Under such conditions (i.e., when $r_{X2,CP} > r_{X2,JP}$) the optimal solutions do contain contextual performance criterion weighting (i.e., the colorful PO curves sometimes outperform the black PO curve; CW+PO outperforms PO alone).

Online Supplement C.

Expanding Results to Hispanic Americans

At the prompting of a helpful reviewer, we also considered the role of the criterion weighting strategy in potentially reducing adverse impact for Hispanic applicants. To achieve this end, we compiled Hispanic-White subgroup differences (i.e., cognitive test d = .72; Roth et al., 2001, p. 318; Conscientiousness d = -.08; Agreeableness d = .05; Foldes et al., 2008, p. 597; structured interview d = .17; Huffcutt & Roth, 1998, p. 186; biodata d = .15; Tenbrink et al. (2025), p. 11; integrity test d = -.05; Dahlke & Sackett, 2017, p. 1409—following the approach of Sackett et al., 2022, we averaged together the effects from overt integrity tests and personality-based integrity tests). The general pattern of Hispanic-White differences on predictors (subgroup d estimates, shown in Table C1) is remarkably similar to the pattern of Black-White differences on the same predictors (shown in Table 2), with two exceptions: Black-White differences are notably larger than Hispanic-White differences for biodata (d = .33 vs. .15) and for integrity tests (d = .10 vs. -.05). We next address the implications of such subgroup differences for criterion weighting effects on PO tradeoff curves.

Thus, we assessed the implications of criterion weighting for Hispanic-White adverse impact under the three scenarios: the Hattrup et al. (1997) 2-predictor scenario, the updated correlation matrix 2-predictor scenario, and the updated correlation matrix 6-predictor scenario. This enabled us to evaluate whether the observed results for Black-White adverse impact reported earlier would also apply to Hispanic-White adverse impact. These results are reported in Figure S1 (for Hattrup et al, 1997, 2-predictors), Figure S2 (for the updated correlation matrix, 2-predictors), and Figure S3 (for the updated correlation matrix, 6-predictors) in the online supplementary material.

First, for the Hattrup et al. (1997) 2-predictor scenario, we used the older correlation matrix from Hattrup et al. (1997). Comparing Figures S1a-b (Hispanic-White tradeoff curves, Hattrup et al. 1997 2-predictor scenario), against Figures 2a-b (Black-White tradeoff curves, Hattrup et al. 1997 2-predictor scenario), we find an almost identical pattern of results. That is, Figure S1a (Hispanic-White PO tradeoff curves for different criterion weights) is nearly identical to Figure 2a (Black-White PO tradeoff curves for different criterion weights). Also, Figure S1b (predictor weights for Hispanic-White PO tradeoff curves) and Figure 2b (predictor weights for Black-White PO tradeoff curves) are virtually identical.

Next, by comparing Figures 3a-b (Black-White tradeoff curves, updated 2-predictor scenario) against Figures S2a-b (Hispanic-White tradeoff curves, updated 2-predictor scenario), we again find an almost identical pattern of results. That is, both the PO tradeoff curves and the corresponding predictor weights are roughly the same for Hispanic-White adverse impact as for Black-White adverse impact.

Further, by comparing Figures 4a-b (Black-White tradeoff curves, updated 6-predictor scenario) against Figures S3a-b (Hispanic-White tradeoff curves, updated 6-predictor scenario), we again find an almost identical pattern of results involving the PO tradeoff curves. However, when inspecting the predictor weights corresponding to the Hispanic-White vs. Black-White PO tradeoff curves, there is one difference. Inspecting the Black-White PO tradeoff curve predictor weights (Figure 4b), as we move rightward across the tradeoff curve to the higher-diversity solutions, diversity is achieved by assigning the greatest predictor weight to Conscientiousness. In contrast, for the Hispanic-White PO curve predictor weights (Figure S3b), as we move rightward across the tradeoff curve to the higher-diversity solutions, diversity is achieved by assigning integrity tests the greatest predictor weight, until the CP criterion weights exceed .5,

after which Conscientiousness becomes the highest-weighted predictor on the diversity end of the tradeoff curve. In other words, for optimizing Hispanic-White adverse impact with 6 predictors, integrity tests end up playing a larger role when high criterion weight is assigned to job performance, because the Hispanic-White *d* for integrity tests is roughly as small as the Hispanic-White *d* for Conscientiousness (indeed, for the Hispanic-White predictor subgroup differences, both integrity test and Conscientiousness *d*'s are slightly below zero—slightly favoring Hispanic applicants over White applicants).

Second, we considered all three groups simultaneously (White, Black, and Hispanic), using the three-group Pareto Optimal method from De Corte et al. (2024). These results are shown in Figure S4 (each criterion weighting scheme shown separately) and Figure 5 (each criterion weighting scheme shown altogether). Figure S4 and Figure 5 again convey that the general pattern of PO tradeoff curves across different criterion weights remains similar for Black-White and Hispanic-White tradeoff curves. Taking this analysis a step further (imitating De Corte et al., 2024), we array personnel selection solutions using sets of three outcomes simultaneously (i.e., criterion validity, Black-White adverse impact, and Hispanic-White adverse impact) in Table C2. Table C2 tabulates the "diversity improvement" (Wee et al., 2014), which is the increase in AI ratio that can be obtained via the PO predictor weighting approach, while holding criterion validity constant. In this demonstration, criterion validity is held constant at the level that would be achieved by using unit weights for the predictors. Table C2 reveals two trends.

The first pattern shows, across all levels of criterion weighting (i.e., from JP:CP = 1:0 to 0:1), the magnitude of diversity improvement obtained through PO predictor weighting remains relatively stable and comparable for both Black–White and Hispanic–White comparisons.

Specifically, the diversity improvement (i.e., the increase in AI ratio relative to unit weighting) ranges from .20 to .24 for Black–White and from .20 to .26 for Hispanic–White. This suggests that the three-group PO approach benefits both minority groups to a similar extent. The second pattern shows that, as the criterion weight assigned to contextual performance increases from zero to 1.0, diversity-related outcomes remain largely unaffected. While criterion validity under unit weighting declines substantially (from .51 to .34), the adverse impact ratios remain constant—approximately .53 for Black–White and .66 for Hispanic–White—regardless of the weight placed on contextual performance. More importantly, the magnitude of diversity improvement from using the PO method, relative to unit weighting, stays consistent across all criterion weighting schemes for both subgroup comparisons. That is, increasing the criterion weight on contextual performance does not yield any meaningful incremental improvement in the AI ratio beyond the gains already achieved via PO predictor weighting. This aligns with our main conclusion: criterion weighting offers little to no benefit for diversity enhancement beyond the PO approach.

Table C1

Hispanic–White and Black–White Group Differences

	Variable	Black–White *d*[a]	Hispanic–White *d*
1	Cognitive Ability	.79	.72[b]
2	Conscientiousness	-.07	-.08[c]
3	Structured Interview	.23	.17[d]
4	Biodata	.33	.15[e]
5	Agreeableness	.03	.05[c]
6	Integrity Test	.10	-.05[f]

Note. [a] Sackett et al., 2022; [b] Roth et al., 2001, Table 7, p. 318; [c] Foldes et al., 2008, Table 4, p. 597; [d] Huffcutt & Roth, 1998, Table 3, p. 186; [e] Tenbrink et al. (2025), Table 10, p. 11; [f] Dahlke & Sackett, 2017, p. 1409, following the approach of Sackett et al., 2022, we averaged together the effects from overt integrity tests and personality-based integrity tests. For all Hispanic-White *d* values, the cited authors considered but decided not to correct for range restriction, with the exception of the Hispanic-White *d* for structured interview, for which no mention was made of range restriction.

Table C2

AI Ratio "Diversity Improvement" due to using PO Weighting instead of Unit Weighting at each Level of Relative Criterion Weighting

CW JP:CP Ratio	Criterion Validity (both for predictor Unit Weighting and PO Weighting)	Black-White AI Ratio (predictor Unit Weighting)	Black-White AI Ratio (predictor PO Weighting)	Black-White Diversity Improvement due to predictor PO Weighting	Hispanic-White AI Ratio (predictor Unit Weighting)	Hispanic-White AI Ratio (predictor PO Weighting)	Hispanic-White Diversity Improvement due to predictor PO Weighting
1:0	0.51	0.53	0.77	0.24	0.66	0.92	0.26
0.83:0.17	0.52	0.53	0.75	0.22	0.66	0.91	0.25
0.75:0.25	0.51	0.53	0.76	0.23	0.66	0.92	0.26
0.5:0.5	0.48	0.53	0.75	0.22	0.66	0.90	0.24
0.25:0.75	0.42	0.53	0.73	0.20	0.66	0.88	0.22
0.17:0.83	0.39	0.53	0.74	0.21	0.66	0.87	0.21
0:1	0.34	0.53	0.73	0.20	0.66	0.86	0.20

Note. CW = criterion weighting; JP = job performance; CP = contextual performance; PO = Pareto-optimal.

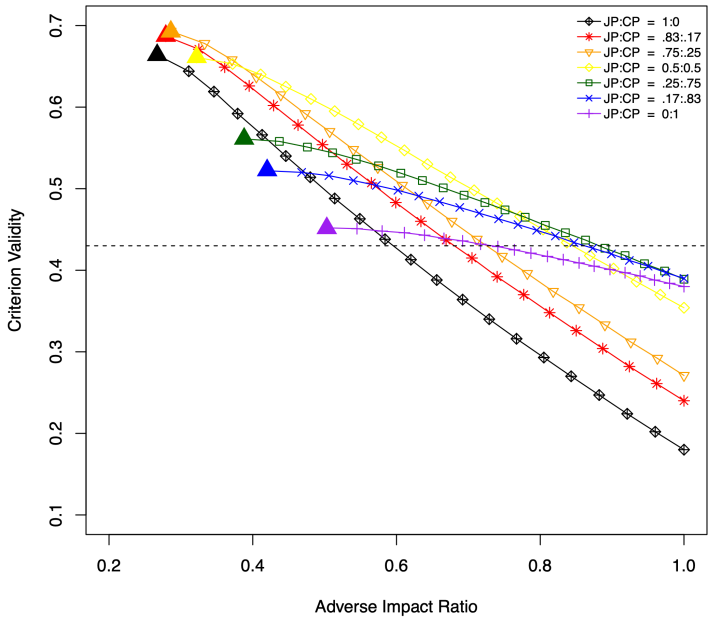


Figure C1a
[Hispanic-White] Pareto-Optimal Tradeoff Curves Based upon Hattrup et al. (1997) Matrix with only Two Predictors

Note. Hattrup et al. (1997) used d = 1.0 to represent the Black-White subgroup difference in cognitive tests, which is similar to Roth et al.'s (2011, p. 311, Black-White d = .99 for industrial samples, N = 464,201). To be consistent with the 1997 correlation matrix, for the Hattrup et al. (1997) scenario, we inserted d = .83 for the Hispanic-White subgroup difference in cognitive tests in industrial samples (Roth et al., 2001, p. 318, N = 313,635), which was the best available estimate for Hispanic-White d that was parallel to the Black-White d used by Hattrup et al. in 1997.

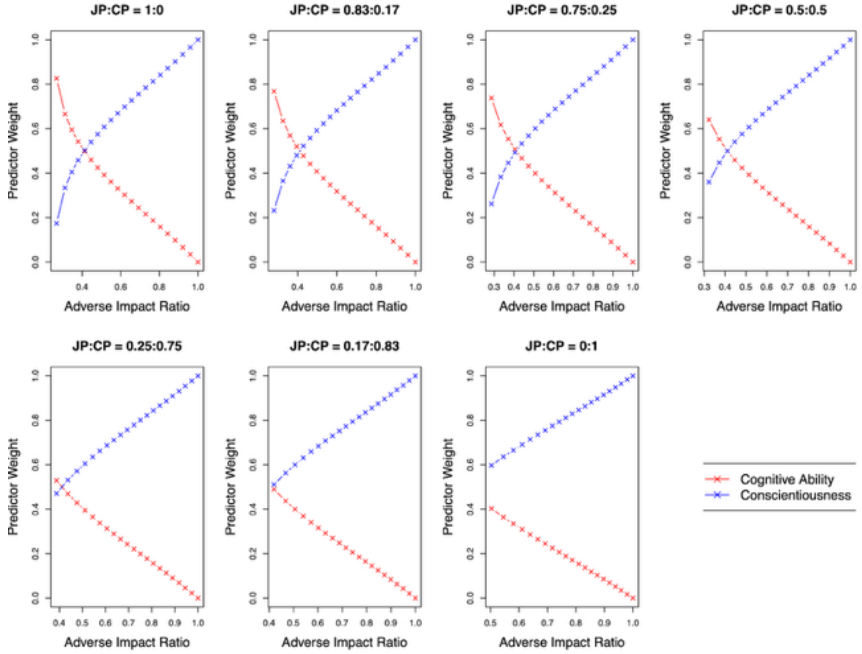


Figure C1b
[Hispanic-White] Predictor Weights Based upon Hattrup et al., (1997) Correlation Matrix

Note. JP = job performance; CP = contextual performance.

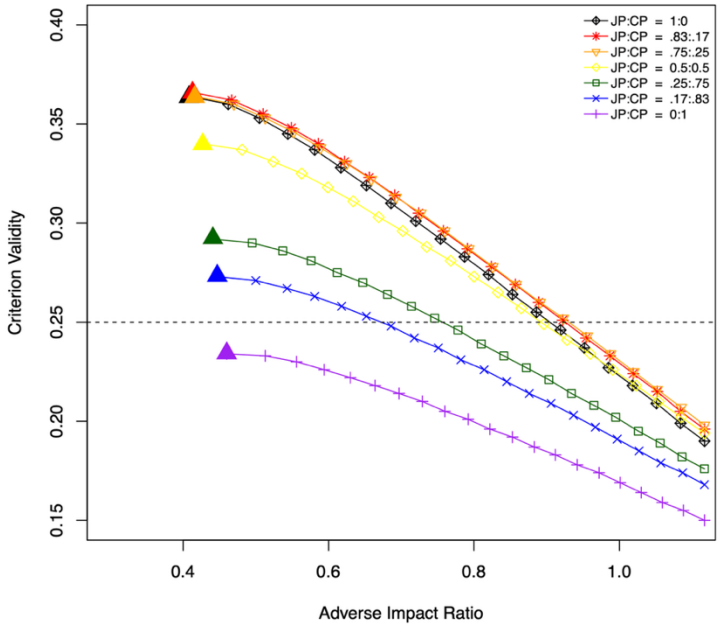


Figure C2a
[Hispanic-White] Pareto-Optimal Tradeoff Curves Based upon Updated Correlation Matrix with only Two Predictors

Note. JP = job performance; CP = contextual performance.

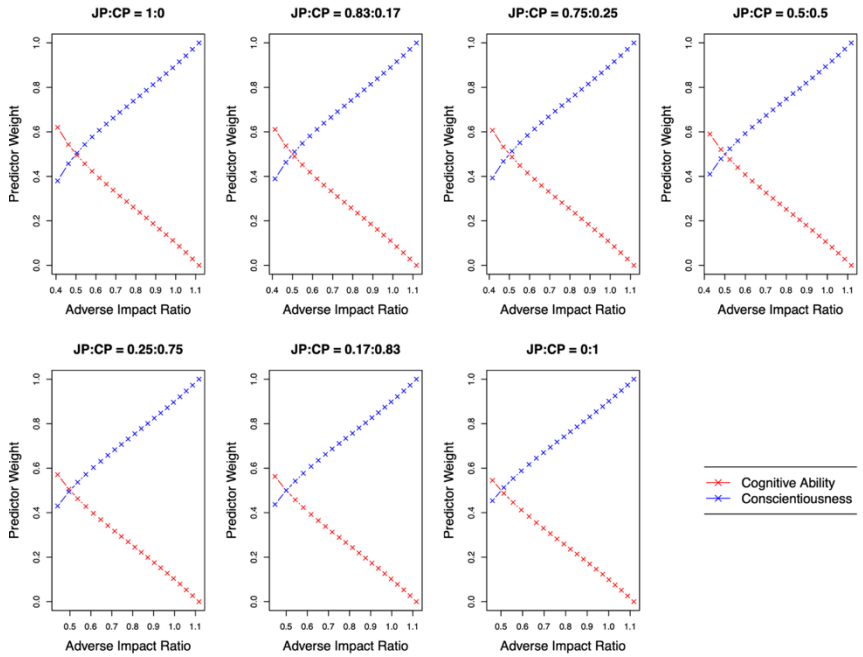


Figure C2b
[Hispanic-White] Predictor Weights Based upon Updated Correlation Matrix with only Two Predictors

Note. JP = job performance; CP = contextual performance.

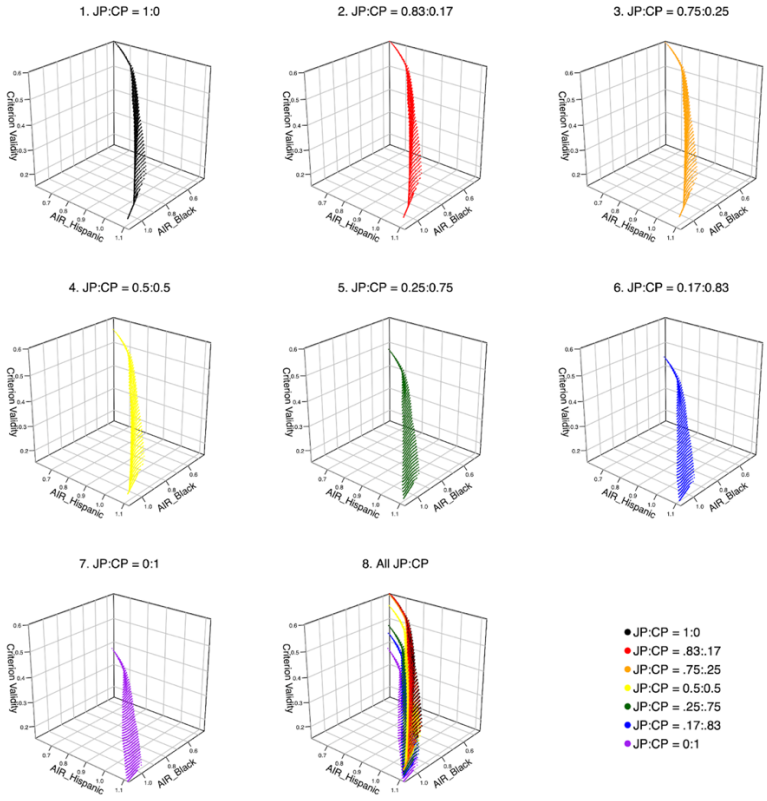


Figure C3

[Hispanic-Black-White] Pareto Front for the Three-Objective Selection Situation

Note. JP = job performance; CP = contextual performance.