Received: 8 January 2024 | Revised: 20 September 2024 | Accepted: 24 September 2024

DOI: 10.1111/ijsa.12501

RESEARCH ARTICLE

International Journal of Selection and Assessment | WILEY

Ethnic-based score differences and differential prediction of various cognitive and noncognitive admissions instruments in higher education

Pieter M. van Lamoen[1] | Annemarie M. F. Hiemstra[1] | Marieke Meeuwisse[1] | Lidia R. Arends[1,2] | Sabine E. Severiens[1,3]

[1]Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

[2]Department of Epidemiology & Biostatistics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

[3]Department of Education for Primary Education, Erasmus University Rotterdam, Rotterdam, The Netherlands

Correspondence
Pieter M. van Lamoen, Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands.
Email: vanlamoen@essb.eur.nl

Abstract

Selective admissions in higher education based on preuniversity grade point average (pu-GPA) can disadvantage ethnic minority students' admission chances. In this study, assessments of motivation and a curriculum-sample test (CST) were compared with pu-GPA by evaluating their effectiveness in mitigating ethnic-based score differences while maintaining predictive validity over course grades. Data obtained from students ($N = 306$) in a high-stakes admissions context were used. The motivation assessments and CST revealed no ethnic-based score differences. Pu-GPA and the CST were found to be positive predictors of the achievement of ethnic majority and ethnic minority students. The motivation assessments were poorly related to achievement. We conclude that CSTs show promise as a method for reducing ethnic-based score differences in admissions procedures while maintaining predictive validity.

KEYWORDS

differential prediction, ethnic-based score differences, higher education, predictive validity, selective admissions

Practitioner Points

- Preuniversity grades are valid predictors of academic achievement in higher education (HE).
- Using preuniversity grades for selection reduces ethnic minority students' chances of admission due to subgroup score differences in grades.
- No ethnic-based score differences were found on assessments of motivation and a curriculum-sample test (CST).
- Preuniversity grades and the CST predicted the academic achievement of ethnic majority and ethnic minority students, but motivation did not.
- HE selective admissions procedures must be rigorously evaluated to ensure equal opportunities for all students.

- CSTs, compared with preuniversity grades, show promise in lowering ethnic-based score differences while maintaining predictive validity.

In response to high numbers of applicants and limited availability of enrolment slots, higher education institutions (HEIs) implement selective admissions procedures. With these procedures, HEIs often aim to admit well-performing and diverse student populations, contribute to equitable opportunities in education, and better equip students to serve increasingly diverse societies (Bowman, 2010; Gurin et al., 2002). Reaching this aim, however, remains a challenge. For instance, in the Netherlands, where the current study is conducted, research indicates that ethnic minority students, defined as students with at least one parent born outside of the Netherlands, have a lower chance of admission into selective programmes (Mulder et al., 2023; Stegers-Jager, 2018). Contributing to this disparity is the common practice of selecting students based on their preuniversity academic achievement, which is typically measured by their preuniversity grade point average (pu-GPA). Although pu-GPA is highly predictive of student achievement in higher education (HE), it also reduces the selection probabilities for ethnic minority students (Richardson et al., 2012; Rosinger et al., 2021; Stegers-Jager et al., 2015).

Admission decisions of HEIs are mostly not only based on preuniversity grades, however. Typically, composite scores are used that combine grades with other types of admissions instruments, such as knowledge tests and assessments of motivation (Fikrat-Wevers et al., 2023; Kurysheva et al., 2019). In the Dutch HE context, 70% of selective HE course programmes use assessments of motivation to select students, and approximately 30% of selective programmes select students on the basis of content-specific exams (Dutch Inspectorate of Education, 2023). The use of these instruments could improve ethnic minority students' selection probabilities, as these instruments may exhibit smaller ethnic-based score differences as compared to pu-GPA (Bradburn & Schmitt, 2019; Ployhart & Holtz, 2008). Combing pu-GPA with different non-grade admissions instruments has therefore been suggested to attract a diverse student population in HE.

Given the high stakes involved in the decisions made in selective admissions, it is important to consider not only how non-grade admissions instruments affect ethnic-based score differences, but also whether these instruments meet other requirements of predictive validity and fairness. Predictive validity usually refers to the relation between test scores and relevant outcomes, such as academic performance (Berry, 2015). What constitutes fairness in selective admissions is contested, but there is consensus that absence of differential prediction is a key requirement (Zwick, 2019). Differential prediction refers to differences between subgroups in regression equations predicting outcomes from admission test scores, which is indicated by differences in regression slopes and/or intercepts (Berry, 2015). Differential prediction is considered particularly problematic when the performance of underrepresented groups is underpredicted (Kennet-Cohen et al., 2014; Shewach et al., 2017).

Despite the wide use of instruments such as assessments of motivation and knowledge tests in the selective admissions practice in HE, research on these types of instruments lags behind research on pu-GPA. In particular, it is unclear how these instruments perform on the aforementioned key indicators of validity and fairness, and specifically differential prediction. For motivation assessments, questions on predictive validity and differential prediction have been addressed, but mostly within low-stakes contexts that are difficult to compare with high-stakes admissions contexts (Isik et al., 2018; Niessen & Meijer, 2017). To assess whether these instruments have the potential to lower ethnic-based score differences caused by pu-GPA while maintaining good predictive validity, it is important to evaluate them in an actual high-stakes admissions context. In the current study, we used high-stakes admissions data obtained from a HE selective admissions procedure that utilised a composite score consisting of pu-GPA, assessments of motivation, and a curriculum-sample test (CST), and we examined how these instruments perform in terms of ethnic-based score differences, predictive validity, and differential prediction.

1 | ASSESSMENT OF MOTIVATION

Assessments of motivation belong to a broader category of non-cognitive admissions instruments that have been proposed to reduce ethnic-based score differences. The idea behind this approach is that the negative effect of pu-GPA on ethnic minority students' selection probabilities may be reduced by adding noncognitive predictors to a composite score, because smaller ethnic sub-group differences are typically found on these predictors (Hough et al., 2001; Schmitt et al., 2009). Research on assessments of motivation remains inconclusive on the question whether this instrument causes ethnic-based score differences. In low-stakes contexts, there are no clear indications of differences in motivation between ethnic groups (Isik et al., 2018), and some studies even suggest that ethnic minority students are more motivated (Hill & Wang, 2015). This might imply that, compared with pu-GPA, motivational assessments show smaller ethnic-based score differences and thus improve ethnic minority students' selection probabilities. By contrast, in high-stakes selection contexts, score differences between ethnic groups may arise, as students from majority groups generally have more access to support when completing assessments of motivation. This may be the case especially when motivation is assessed by personal statements (Woo et al., 2023; Zamanzadeh et al., 2020).

Regarding predictive validity, studies have found that noncognitive predictors, such as motivation, are valuable in predicting academic success and offer incremental value over prior academic performance (Credé & Kuncel, 2008; Richardson et al., 2012; Schmitt et al., 2009). Noncognitive predictors are for instance indicative of the extent to which students' motives and interests fit their programme of choice, which contributes to academic achievement (Soppe et al., 2024). However, the conclusions of these studies are generally based on data collected in low-stakes contexts, with low ecological validity to high-stakes selection contexts (de Boer & van Rijnsoever, 2022). Accurate evaluation of motivation in high-stakes selection contexts is highly challenging and prone to faking, which possibly lowers the predictive validity of assessments of motivation (Niessen & Meijer, 2017; Peterson et al., 2011). Furthermore, regarding differential prediction, research suggests that the relationship between motivation and academic performance may vary between ethnic majority and ethnic minority students. Specifically, the academic performance of ethnic majority students may be better predicted by motivation than the performance of ethnic minority students (Isik et al., 2017). These results highlight a need to examine ethnic-based score differences, predictive validity, and differential prediction of motivation assessments in a high-stakes admissions context.

2 | CST

A second instrument that HEIs increasingly use alongside pu-GPA to select students is the CST. In CSTs, applicants perform tasks similar to those in their future study programme. For example, they study domain-specific materials and take an exam (Lievens & Coetsier, 2002; Niessen et al., 2016). The rationale behind this approach is that, following the theory of behavioural consistency (Wernimont & Campbell, 1968), a representative sample of the study programme taps into the same cognitive and noncognitive factors that are necessary for achievement within the actual study programme (Niessen et al., 2016).

Multiple studies have found that CSTs can accurately predict academic performance in terms of grades and study progress (de Visser et al., 2018; Niessen et al., 2016; Reibnegger et al., 2010). It has also been suggested that curriculum sampling could help to reduce differential prediction by gender in high-stakes admissions contexts (Niessen et al., 2019). Less research has been conducted on performance differences and differential prediction of CSTs by ethnicity. A recent study in medical education revealed that ethnic minority students scored significantly lower on the CST compared with ethnic majority students (Fikrat-Wevers et al., 2023). Additional research is needed to examine the generalisability of these findings, particularly given the domain-specificity of CSTs. Furthermore, research has, to our knowledge, not yet examined ethnic-based differential prediction of CSTs.

3 | THE CURRENT STUDY

In this study, we aim to contribute to existing literature on the validity and fairness of instruments that are used in HE selective admissions procedures. To this end, we examine several instruments, namely pu-GPA, assessments of motivation, and a CST, that were used in the context of a high-stakes admissions procedure of a psychology programme in the Netherlands. The combined use of these different types of admissions instruments in a single admissions procedure provides a unique opportunity to evaluate and compare these instruments on various key indicators of validity and fairness. We aim to do this by addressing three research questions. Firstly, we ask whether, and to what extent, there are ethnic-based score differences on the admission instruments pu-GPA, assessments of motivation, and the CST, and to what extent the combined usage of these instruments affects the ethnic diversity of a student population. Secondly, we ask to what extent pu-GPA, assessments of motivation, and the CST are valid predictors of students' academic performance, measured by course grades in HE. Thirdly, we ask to what extent these admissions instruments show ethnic-based differential prediction of students' course grades.

4 | METHOD

4.1 | Procedure and sample

Data were collected in 2020 at an undergraduate psychology programme of a research university in an urban area of the Netherlands. Since that year, applicants to the programme have been required to complete an admission assessment, consisting of four components: pu-GPA, a CST, strength of motivation for studying psychology (PSY-motivation), and study motivation. Applicants completed the CST and the motivation assessments via an online tool that could be filled out at any location. Applicants prepared for the CST by reading chapters and watching video lectures. The assessment took approximately 7 h to complete, including 5 h of preparation. In total, 1361 applicants completed the assessment. Applicants' scores on the four separate components were z-transformed and combined into a composite score with each component weighted at 25%.

The composite score was used to allocate 600 available enrolment slots. Eventually, 562 students enrolled in the study programme, because not all admitted applicants accepted their invitation to enrol. International students were excluded from the analyses, because our study concerns differences between Dutch ethnic majority and ethnic minority students. The sample used for our analyses included 306 first-year students (M_{age} = 19.45, SD_{age} = 2.40, 83.4% female) who applied for the undergraduate psychology programme in 2020 and enrolled in the programme in the 2020–2021 academic year.

We obtained achievement data (i.e., course grade and study progress) and student background data through the university's

educational research database and linked that information to admissions procedure data by using student identification numbers. The procedures of this study were approved by the department's ethical committee (reference number 21-019).

4.2 | Measures

4.2.1 | Preuniversity academic achievement

Preuniversity academic achievement was operationalised as pu-GPA. Applicants were asked to submit their grades for either the penultimate year of prior education or the last year if that year had already been completed. Most students (i.e., approximately 70%) submitted grades from their fifth year in secondary education, as most students enrol in the Dutch-language psychology programme after finishing preuniversity education, which is a 6-year educational track.

4.2.2 | Assessments of motivation

The assessment of applicants' PSY-motivation consisted of three parts. First, applicants had to write short personal statements with a maximum of 150 words to answer two open questions: "Why do you want to study psychology?" and "Why do you want to study psychology at our university?" Second, applicants were asked which activities they had undertaken to obtain information about the psychology programme (e.g., visiting the open day or reading the brochure of the programme) and which extracurricular activities made them a good match for the programme (e.g., a side job or voluntary work). Third, applicants filled in a questionnaire aimed to measure their strength of motivation to study psychology. The 15-item questionnaire was developed and used in the medical education context and was found to be a reliable and valid instrument to measure strength of motivation for this specific field of study (Nieuwhof et al., 2004; Wouters et al., 2016). Items were adapted to the psychology context. An example item is "I would always regret not attempting to study psychology." Answer categories for all questions ranged from 1 (completely disagree) to 5 (completely agree). The 15-item questionnaire had an estimated reliability of $\alpha = .75$. The PSY-motivation sub-assessments were weighted as follows: personal statements (40%), attended activities (40%), questionnaire (20%).

Study motivation was assessed via a 12-item questionnaire that was developed in the medical education context to measure autonomous and controlled study motivation (Wouters et al., 2016). An example item is "I will participate actively in the psychology courses because I feel like it is a good way to improve my understanding of the material." Applicants indicated the extent to which the statements applied to them on a 7-point scale (1 = not true at all; 7 = very true). The estimated reliability of the scale was $\alpha = .60$.

4.2.3 | CST

The CST was designed to realistically represent the first part of the undergraduate programme in psychology. The materials provided to students by the HEI were therefore at an undergraduate year-1 level, and the online lectures were provided by the psychology programme's lecturers, who also developed the exam questions. For each of the three subjects (personality, biological, and clinical psychology), five multiple-choice questions were included in the CST, each with four answer options. Two questions about the subject of clinical psychology involved performing statistical calculations, with the aim of also representing the methods and statistics component of the psychology programme. The CST's estimated reliability was $\alpha = .44$, signifying low reliability since an α score of .70 is typically the minimum required to indicate adequate reliability (Nunnally & Bernstein, 1994). This low reliability can be explained by the heterogeneous content of the test (Lievens & Coetsier, 2002), which included questions concerning three subfields of psychology and statistics, and by the low number of items that were included in the CST. Given the heterogeneous content of the test, test–retest reliability would provide a better estimate of the CST's reliability; however, this was not possible due to the absence of repeated measures. As an approximation of test–retest reliability, we computed a coefficient of split-half reliability by using the Spearman–Brown formula (for an explanation, see Eisinga et al., 2013). This led to a reliability coefficient of 0.51. Given the alignment of the CST with the content coverage and learning objectives of the psychology programme, we assume the CST to be a valid indicator of applicants' content knowledge after they studied the materials provided to them.

4.2.4 | Academic success

Academic success, the outcome or criterion measure of this study, was defined as achievement (i.e., grades) and study progress (i.e., credits).

Academic achievement was measured by first-course grade and first-year grade point average (FYGPA). Grades for separate courses were given on a scale from 1 (lowest) to 10 (highest), but only grades above 4 were registered in the university's educational research database. First-course grade is the final grade (i.e., the grade after possible resits) that students received for the first course in the psychology programme, an introductory course in social psychology. In this course, 97.1% of the students obtained a grade of 4.0 or higher. FYGPA was calculated as the average of the final grades for all courses, weighted by the credit hours (European credit transfer and accumulation system, ECTS) assigned to these courses. We did so for all students who had obtained at least one grade for a first-year psychology course (cf. Kickert et al., 2021).

Study progress of the first year was operationalised as obtaining the required number of ECTS credits to enrol in Year 2 of the psychology programme. For cohorts 2018 and 2019 (which

were included in our study to analyse potential inequalities based on the 2020 selective admissions procedure), this meant a total of 60 (out of 60) ECTS. For the 2020 cohort, the programme management decided, because of the COVID-19 pandemic, students had to obtain 52.5 ECTS to enrol in Year 2. First-year study progress was a binary variable (successful/not successful).

4.2.5 | Ethnic background

We retrieved ethnic background data from the university's educational research database. The distinction between ethnic majority and ethnic minority students was based on the definition used by Statistics Netherlands. According to this definition, an individual belongs to an ethnic minority group if at least one parent was born outside the Netherlands. Based on their parents' countries of birth, ethnic minority students were classified as having a non-Western background, including primarily a Moroccan, Surinam, Turkish, or Antillean background. Research has shown that non-Western ethnic minority students in particular experience disadvantages in HE, compared to Western ethnic majority (Dutch) students and other Western ethnic minority students (e.g., from Western European countries) (Stegers-Jager et al., 2015). To answer our research questions and due to the limited number of students with a Western minority background (n = 39), we decided to exclude those students from the analyses.

4.3 | Analyses

4.3.1 | Predictive validity

Predictive validity refers to the extent to which predictors and outcome(s) correlate with each other. We assessed this by calculating the Pearson correlations of predictors and outcomes, both uncorrected and corrected for restriction of range. In selection research, restriction of range is a common problem in analyses, because of (educational) organisations' preference for high-ability applicants implies that admitted applicants have a higher ability and are more similar in ability than the complete applicant group (Dahlke et al., 2019). In the current study, students were selected on an operational composite of the various predictors, leading to indirect range restriction on each individual predictor, which leads to an underestimation of the validity of the admissions instruments (Sackett & Yang, 2000; Zimmermann et al., 2017). Estimates of the predictive validity of the separate predictors were therefore corrected for range restriction by using the Pearson-Lawley multivariate correction method (Lawley, 1944). Estimates of the predictive validity of the composite score used for selection were corrected by using Thorndike's case 2 correction formula for direct range restriction, as direct selection occurred on this composite score (Thorndike, 1949).

4.3.2 | Ethnic-Based score differences

To assess differences between groups from different ethnic backgrounds on the predictors, we adopted two complementary approaches. First, we calculated sub-score differences between ethnic majority and ethnic minority students on the predictors and the final composite, tested them for significance with t-tests, and expressed them in terms of Cohen's d effect sizes. Ethnic background information was only available for students who had enrolled in the programme, not for applicants who had not enrolled, leading to restriction of range in this analysis. Therefore, to avoid underestimation of sub-score group differences, we calculated d-values based on the range restriction corrected correlation matrix (Wiernik & Dahlke, 2020). Second, we examined ethnic-based score differences at a student population level by comparing the ethnic diversity of the 2020-cohort, who had enrolled with selective admissions, with the ethnic diversity of the 2018- and 2019-cohorts, who had enrolled in the same study programme through open admissions.

4.3.3 | Differential prediction

Differential prediction refers to differences in regression equations of predictors on outcomes between subgroups (Berry, 2015). To examine differential prediction, we used the Cleary model of test bias (Cleary, 1968) by conducting step-down hierarchical regression analyses (Aguinis et al., 2010; Lautenschlager & Mendoza, 1986). In line with this procedure, we first estimated three regression models for each combination of predictor and outcome: Model 1 included only the predictor; Model 2 included the predictor and the categorical ethnicity variable; and Model 3 included the predictor, ethnicity, and the interaction between the predictor and ethnicity. Then, to formally assess whether differential prediction existed, we conducted an omnibus test to compare Models 1 and 3. If Model 3 showed incremental explained variance over Model 1 ($\Delta R^2_{\text{Omnibus,}}$), then evidence of differential prediction existed. In the next step, we tested whether the differential prediction was due to differences between slopes, intercepts, or both. We examined slope differences by comparing Model 3 against Model 2. Here, increased explained variance ($\Delta R^2_{\text{Slope}}$) indicated slope differences between ethnic groups. Lastly, we assessed intercept difference by comparing Model 2 against Model 1, with increased explained variance ($\Delta R^2_{\text{Intercept}}$) indicating intercept differences between ethnic groups.

Before conducting regression analyses, we applied corrections for range restriction to covariance matrices including the predictors, outcomes, ethnic group variable, and interaction variables (predictor*ethnicity). These covariance matrices were used as input for the regression analyses. The observed sample size (n = 306) was used to evaluate the significance tests for the Cleary test analyses.

We further analysed potential slope and intercept differences to determine whether over- or underprediction occurred for the minority group relative to the majority group. In addition to examination of regression coefficients and ΔR^2, we calculated

categorical moderation (d_{Mod}) effect sizes to quantify the magnitude of over- or underprediction (Dahlke & Sackett, 2018; Nye & Sackett, 2017). The following d_{Mod} effect sizes were computed: d_{Mod_Signed} (signed effect size for differential prediction, with positive values when ethnic minority students' performance was overpredicted and negative values when ethnic minority students' performance was underpredicted), $d_{Mod\_Unsigned}$ (unsigned effect size for differential prediction, regardless of whether it took the form of over- or underprediction), d_{Mod_Under} (standardised difference in the score range where underprediction for ethnic minority students occurred), and d_{Mod_Over} (standardised difference in the score range where overprediction for ethnic minority students occurred). The d_{Mod} effect sizes can be interpreted similarly to Cohen's *d*. The *psychmeta* package in R was used to compute the effect sizes.

5 | RESULTS

Table 1 presents descriptive statistics and correlations of this study's variables, including the four predictors (pu-GPA, the CST, PSY-motivation, and study motivation), the composite score, the ethnic grouping variable, and the outcome variables first-course grade and first-year GPA. The PSY-motivation and composite scores were both constructed from *z*-scores and thus have means around zero. For the four predictors and the composite score, the mean and standard deviation of both the selected sample (i.e., students enrolled in the course programme; restricted sample) and the applicant sample (unrestricted sample) are presented together with the predictors' *u*-ratios (restricted predictor SD divided by unrestricted predictor SD), indicating the amount of range restriction of these predictors.

5.1 | Predictive validity

We used the correlations between predictors and outcomes in Table 1 to examine predictive validity. As listed in this table, we found moderate positive and significant ($p < .001$) correlations between pu-GPA on the one hand and first-course grade and FYGPA on the other, with *r*-coefficients corrected for range restriction of 0.29 (first-course grade) and 0.50 (first-year GPA), implying that students who earned higher grades in preuniversity education also obtained higher grades in their first year of HE. Among the various predictors, pu-GPA was the strongest predictor of FYGPA. Table 1 further shows significant ($p < .001$) positive (corrected for range restriction) correlations between the CST on the one hand and first-course grade ($r = 0.36$) and FYGPA ($r = 0.36$) on the other. The third predictor, PSY-motivation, revealed positive but nonsignificant correlations between the two outcome variables, although *r*-coefficients increased to values of 0.20 (first-course grade) and 0.22 (FYGPA) after range restriction corrections were applied to the correlation matrix. For the fourth predictor, study motivation, Table 1 indicates significant ($p < .05$) negative correlations of $r = -0.12$ for first-course grade and $r = -0.11$ for FYGPA. However, the two corrected correlations were around zero, indicating no association between study motivation and the two outcomes.

Lastly, regarding the composite score, Table 1 shows positive and significant ($p < .001$) corrected correlations of $r = 0.31$ for first-course grade and $r = 0.43$ for FYGPA. Students with higher scores in the admissions procedure thus performed better during their first year in HE. Furthermore, according to the results in Table 1, the CST has stronger predictive validity for first-course grade than the composite score, and pu-GPA has stronger predictive validity for FYGPA than the composite score.

5.2 | Ethnic-based score differences

Tables 2 and 3 display the results of the analyses that were conducted to examine score differences on the predictors and enrolment differences between ethnic majority and ethnic minority students. In Table 2, the mean scores on the predictors and the composite score are presented for ethnic majority and ethnic minority students. Results of the independent sample t-tests including corrected and uncorrected effect sizes (Cohen's *d*) are also presented in Table 2. Range restriction corrected *d*-values are calculated based on the corrected correlation coefficients of the predictors with the ethnic grouping variable in Table 1 (Wiernik & Dahlke, 2020).

Table 2 shows that ethnic majority students scored higher on pu-GPA ($M = 6.87$, $SD = 0.45$) than ethnic minority students ($M = 6.69$, $SD = 0.50$). This difference is significant, $t(129) = 2.91$, $p = .004$, with a medium corrected effect size of .44. Ethnic majority students scored similarly to ethnic minority students on the CST ($M = 12.87$, $SD = 1.47$ and $M = 12.61$, $SD = 1.36$, respectively), although there was a slight nonsignificant difference, $t(151) = 1.47$, $p = .145$, with an observed effect size of $d = .18$. After we corrected for range restriction, the effect size increased to $d = .25$. Although still low in value, this effect size indicated that there may have been small ethnic-based group differences in the complete unrestricted applicant sample on the CST. On the third and fourth predictors, PSY-motivation and study motivation, no differences were observed – scores were somewhat higher but nonsignificant, and corrected effects sizes were low ($d = -.16$ for PSY-motivation and $d = -0.05$ for study motivation). Lastly, according to Table 2 scores on the composite measure used for selection were similar for ethnic majority students ($M = 0.02$, $SD = 0.47$) and ethnic minority students ($M = -0.05$, $SD = 0.36$); the observed difference was not significant and a low corrected effect size of $d = .20$. was found.

To examine the effect of the selection procedure on the student population's ethnic diversity, we analysed two cohorts enrolled without selective admissions (cohorts 2018 and 2019) and one cohort enrolled with selective admissions (cohort 2020). Table 3 displays the ethnic composition of these cohorts both at the start and at the end of the first year. This table indicates that at the start of the 2020 academic year, 26.47% of the students who had enrolled in the psychology programme had an ethnic minority background, as compared to 30.57% in 2018 and 30.80% in 2019. The percentage of

TABLE 1 Descriptive statistics and intercorrelations among study variables.

		Restricted sample		Unrestricted sample									
		M	SD	*M*	SD	*u*-ratio	1	2	3	4	5	6	7
	Predictors												
1.	pu-GPA	6.82	0.47	6.75	0.52	0.91	-						
2.	Curriculum-sample	12.80	1.42	12.16	1.83	0.78	0.08	-					
3.	PSY-motivation	0.00	0.46	0.02	0.52	0.89	−0.03	0.04	-				
4.	Study motivation	5.12	0.49	4.97	0.55	0.91	−0.17***	−0.22***	−0.13*	-			
5.	Final composite	0.00	0.44	−0.06	0.53	0.84	0.50***	0.51***	0.50***	0.27***	-		
	Background												
6.	Ethnicity[a]	-	-	-	-	-	−0.17**	−0.08	0.11	0.05	−0.07	-	
	Range restriction correction[b]						−0.19	−0.11	0.07	−0.02	−0.09		
	Outcomes												
7.	First-course grade	7.10	1.15	-	-	-	0.23***	0.26***	0.10	−0.12*	0.26***	−0.07	-
	Range restriction correction[b]						0.29	0.36	0.20	−0.02	0.31		
8.	First-year GPA	6.98	0.74	-	-	-	0.44***	0.23***	0.09	−0.11*	0.36***	−0.20**	0.68***
	Range restriction correction[b]						0.50	0.36	0.22	0.01	0.43		

[a]Ethnic majority students are coded as value 0; ethnic minority students as value 1.

[b]Correlations between predictors and background/outcome variables are corrected for multivariate indirect range restriction (Pearson-Lawley). Correlations between the final composite score and background/outcome variables are corrected for direct range restriction (Thorndike Case II). Statistical significance was determined before range restriction corrections.

*$p < .05$;

**<.01;

***<.001.

TABLE 2 Means and standard deviations of admission exam tests and scores for ethnic majority and ethnic minority students (2020-sample).

Predictor	Ethnic majority students (n = 225)		Ethnic minority students (n = 81)		t-value	Observed d-value	Corrected d-value
	M	SD	M	SD			
pu-GPA	6.87	0.45	6.69	0.50	2.91**	0.40	0.44
CST	12.87	1.47	12.61	1.36	1.47	0.18	0.25
PSY-motivation	−0.03	0.48	0.08	0.37	−2.16	−0.25	−0.16
Study motivation	5.0	0.49	5.11	0.50	−0.15	−0.02	−0.05
Composite	0.02	0.47	−0.05	0.36	1.43	0.17	0.20

Note: Positive effect sizes reflect differences favouring ethnic majority students, negative effect sizes reflect differences favouring ethnic minority students.

**p < .01.

TABLE 3 Ethnic group composition of cohorts with and without selective enrolment.

	2018-sample (nonselective)		2019-sample (nonselective)		2020-sample (selective)	
	N	%	N	%	N	%
Enrolled (start Year 1)						
Ethnic majority students	343	69.43	492	69.20	225	73.53
Ethnic minority students	151	30.57	219	30.80	81	26.47
Total	494		711		306	
Successful completion (end Year 1)						
Ethnic majority students	190	78.88	314	76.03	199	75.95
Ethnic minority students	51	21.16	99	23.97	63	24.06
Total	241		413		262	

ethnic minority students was thus lower in the cohort that had enrolled with selective admissions as compared to the cohort that had enrolled without selective admissions. However, the percentage of ethnic minority students who had successfully completed the first year relative to ethnic majority students in the selective 2020 cohort (24.06%) was similar to that of the nonselective 2018 and 2019 cohorts (21.16% and 23.97%, respectively).

5.3 | Differential prediction

Table 4 lists the results for the step-down regression analyses that we used to detect differential prediction. Corresponding d_{Mod} effect sizes are presented in Table 5. Analyses of the four individual predictors were performed using covariance matrices corrected for multivariate indirect range restriction by applying the Pearson–Lawley correction formula. The analysis of the composite score made use of a covariance matrix that was corrected for direct range restriction via the Thorndike Case 2 correction formula. Figure 1 depicts the regression lines for ethnic majority and ethnic minority students.

5.3.1 | Preuniversity GPA

Differential prediction with an intercept difference was found in the model for FYGPA. This was indicated by the significant omnibus comparison and a significant intercept comparison (see Table 4). The nonsignificant slope comparison implies that slopes of regression lines were equal for ethnic majority and ethnic minority students. Note that the signed (d_{Mod_signed} = 0.26) and unsigned ($d_{Mod\_unsigned}$ = 0.33) effect sizes were close in value, which implies that effects did not cancel each other out due to slope differences (Nye & Sackett, 2017). Ethnicity coefficients were positive, meaning that performance of the ethnic minority group was overpredicted. Corresponding effect sizes in Table 5 indicate that this overprediction was of small to moderate magnitude (d_{Mod_over} = 0.30). For the second outcome variable first-course grade no statistically significant differential prediction was found, as indicated by the nonsignificant omnibus test. Note however that the trends in Figure 1 suggest a slope difference that this study could not detect at the α = .05 level due to lack of power.

TABLE 4 Regression analyses of selection predictors on first-year GPA and first-course grade.

Outcomes		Individual predictors				
		pu_GPA	Curriculum-sample	Psy-motivation	Study motivation	Composite
First-year GPA	*Model 1*					
	Predictor	0.76 (0.08)***	0.15 (0.02)***	0.33 (0.09)***	0.02 (0.08)	0.69 (0.09)***
	R^2	0.252	0.126	0.047	0.000	0.173
	Model 2					
	Predictor	0.73 (0.08)***	0.14 (0.02)***	0.36 (0.08)***	0.02 (0.08)	0.67 (0.09)***
	Ethnicity[a]	−0.22 (0.09)*	−0.31 (0.09)***	−0.41 (0.10)***	−0.38 (0.10)***	−0.27 (0.09)**
	R^2	0.267	0.158	0.100	0.046	0.199
	Model 3					
	Predictor	0.78 (0.09)***	0.16 (0.03)***	0.40 (0.09)***	0.03 (0.09)	0.72 (0.10)***
	Ethnicity[a]	−0.23 (0.09)*	−0.32 (0.09)***	−0.38 (0.10)***	−0.38 (0.09)***	−0.28 (0.09)**
	Interaction	−0.16 (0.18)	−0.05 (0.07)	−0.31 (0.24)	−0.04 (0.20)	−0.25 (0.23)
	R^2	0.269	0.159	0.105	0.046	0.203
	Step-down comparison					
	ΔR^2 Omnibus	0.017*	0.033**	0.058***	0.046***	0.030**
	ΔR^2 Slope	0.002	0.001	0.005	0.000	0.004
	ΔR^2 Intercept	0.015*	0.032***	0.053***	0.046***	0.026**
First-course grade	ΔR^2 Intercept					
	Predictor	0.66 (0.13)***	0.24 (0.04)***	0.47 (0.13)***	−0.05 (0.12)	0.79 (0.14)***
	R^2	0.084	0.131	0.041	0.000	0.094
	Step 2					
	Predictor	0.65 (0.13)***	0.23 (0.04)***	0.48 (0.13)***	−0.05 (0.12)	0.78 (0.14)***
	Ethnicity[a]	−0.10 (0.15)	−0.13 (0.14)	−0.28 (0.15)	−0.24 (0.15)	−0.12 (0.14)
	R^2	0.085	0.133	0.051	0.008	0.096
	Step 3					
	Predictor	0.80 (0.16)***	0.21 (0.04)***	0.61 (0.14)***	−0.03 (0.07)	0.87 (0.16)***
	Ethnicity[a]	−0.14 (0.15)	−0.12 (0.14)	−0.22 (0.15)	−0.24 (0.15)	−0.13 (0.14)
	Interaction	0.51 (0.30)	0.10 (0.10)	−0.77 (0.38)*	−0.06 (0.30)	−0.42 (0.38)
	R^2	0.0094	0.136	0.064	0.008	0.099
	Step-down comparison					
	ΔR^2 Omnibus	0.010	0.005	0.023*	0.008	0.005
	ΔR^2 Slope	0.009	0.003	0.013*	0.000	0.003
	ΔR^2 Intercept	0.001	0.002	0.010	0.008	0.002

Note: standard errors are between brackets.

[a]Ethnic majority students are coded as value 0; ethnic minority students as value 1.

*<.05;

**<.01;

***<.001.

5.3.2 | CST

A statistically significant intercept difference was found in the model for FYGPA in Table 4. The significant omnibus comparison firstly indicated differential prediction. The nonsignificant slope comparison and significant intercept comparison then indicated that this differential prediction was due to intercept difference and not slope difference. The signed and unsigned effect sizes being positive and equivalent (d_{Mod_Signed} = 0.42 and

TABLE 5 Effect sizes for differential prediction.

Outcome	Predictor	d_{Mod_signed}	$d_{Mod\_unsigned}$	d_{Mod_under}	d_{Mod_over}	Prop.$_{Under}$	Prop.$_{Over}$
First-year GPA	pu-GPA	0.26	0.33	−0.04	0.30	0.22	0.78
	Curriculum sample	0.42	0.42	0.00	0.42	0.00	1.00
	Psy-motivation	0.57	0.57	0.00	0.57	0.00	1.00
	Study motivation	0.53	0.53	0.00	0.53	0.00	1.00
	Composite	0.39	0.39	0.00	0.39	0.00	1.00
First-course grade	pu-GPA	0.06	0.18	−0.06	0.12	0.38	0.62
	Curriculum sample	0.12	0.14	−0.01	0.13	0.12	0.89
	Psy-motivation	0.24	0.28	−0.02	0.26	0.16	0.84
	Study motivation	0.21	0.21	0.00	0.21	0.00	1.00
	Composite	0.13	0.15	−0.01	0.14	0.16	0.84

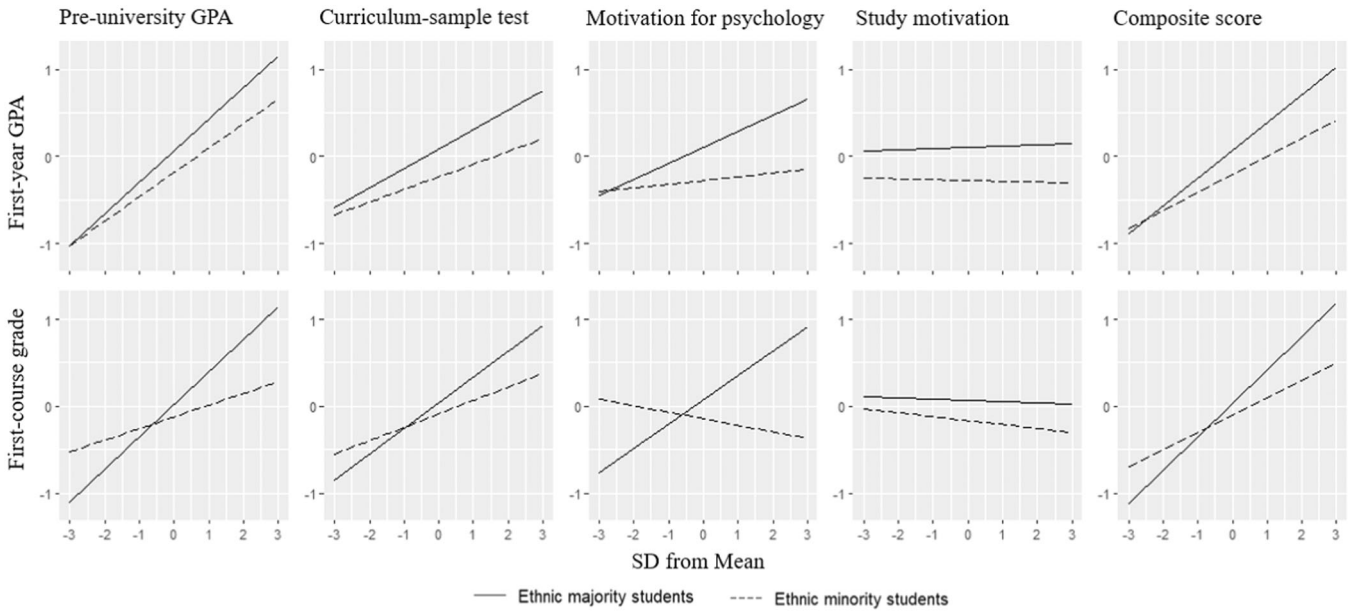

FIGURE 1 Expected FYGPA and First-course grade based on regression models presented in Table 4. FYGPA, first-year grade point average.

$d_{Mod\_unsigned}$ = 0.42) imply that the ethnic minority group's grades were overpredicted, and that this overprediction occurred over the full range of participants' CST scores, which is also visible in Figure 1. No differential prediction was found in the model for first-course grade, as indicated by the nonsignificant omnibus comparison and small effect sizes (d_{Mod_Signed} = 0.12 and $d_{Mod\_unsigned}$ = 0.14).

5.3.3 | PSY-motivation

Statistically significant differential prediction in terms of intercept difference was found in the model for FYGPA. This was indicated by a significant omnibus comparison, a nonsignificant slope comparison, and a significant intercept comparison. Positive and equivalent signed and unsigned effects sizes (d_{Mod_Signed} = 0.53 and $d_{Mod\_unsigned}$ = 0.53) signify that ethnic minority students' grades were overpredicted across the full range of scores. For first course grade, the models showed differential prediction in terms of slope difference, as indicated by a significant omnibus comparison, a significant slope comparison, and a nonsignificant intercept comparison. The interaction plot of PSY-motivation in Figure 1 shows that PSY-motivation was predictive for ethnic majority students' first-course grades, but that it was not predictive for the grades of ethnic minority students.

5.3.4 | Study motivation

As noted above, study motivation was not associated with FYGPA nor with first-course grade. Models in Table 4 further show that this was the case for both ethnic majority and ethnic minority students.

5.3.5 | Composite

Differential prediction in terms of intercept difference was found in the FYGPA models, as indicated by the significant omnibus comparison, nonsignificant slope comparison, and significant intercept comparison in Table 4. Positive and equivalent signed and unsigned effects sizes (d_{Mod_Signed} = 0.39 and $d_{Mod\_unsigned}$ = 0.39) signify that ethnic minority students' grades were overpredicted across the full range of scores. This overprediction was of moderate magnitude and comparable to overprediction caused by the CST. For first course-grade, no statistically significant differential prediction was found, as indicated by the nonsignificant omnibus comparison. Note however that Figure 1 suggests a small slope difference between ethnic majority and ethnic minority students.

6 | DISCUSSION AND CONCLUSION

Selective admissions in HE based on prior educational achievement decrease the probabilities of ethnic minority students to enrol in selective programmes (Fikrat-Wevers et al., 2023; Rosinger et al., 2021). Using admissions instruments other than grades has been suggested to attract a diverse student population in HE course programmes (Stegers-Jager et al., 2015). In practice, many HEIs already select their students based on a wide range of non-grades admissions instruments, but an empirical evaluation of the predictive validity and ethnic-based score differences is lacking (Bradburn & Schmitt, 2019; Niessen & Meijer, 2017). In the present study, we examined whether assessments of motivation and CSTs, two types of instruments that are commonly used in HE selective admissions, show smaller ethnic-based score differences than grades do, and whether these instruments meet requirements regarding predictive validity and ethnic-based differential prediction. To fulfil these objectives, we analysed data obtained in a high-stakes HE admissions procedure that included various admissions instruments: prior academic achievement, a CST, and assessments of PSY-motivation and study motivation.

Our results showed that prior academic achievement was a good predictor of first-year academic success, but also that ethnic minority students' scores on this predictor were lower than those of ethnic majority students. These findings regarding pu-GPA's predictive validity and score differences by ethnic background are consistent with prior research (Rosinger et al., 2021; Stegers-Jager, 2018; Vulperhorst et al., 2018). Our results also support previous research on the strategy to reduce score differences between students from different ethnic backgrounds by adding noncognitive indicators alongside pu-GPA to a final composite score (Bradburn & Schmitt, 2019; Schmitt et al., 2009). Neither the PSY-motivation nor study motivation assessments showed score differences by ethnic background. This absence of score differences on the motivation assessments compensated for the score differences that we observed on pu-GPA in the final composite score, which did not show score differences between ethnic majority and ethnic minority students.

Our results further indicated small but nonsignificant score differences between ethnic majority and ethnic minority students on the CST. These results suggest that CSTs might be a viable alternative to pu-GPA to increase ethnic minority students' chances of admission into HE course programmes, as pu-GPA showed larger ethnic-based differences in scores. These results contrast with those of Fikrat-Wevers et al. (2023), who found that ethnic minority students scored significantly lower on a CST in the medical educational context. However, this effect was not present in all educational programmes included in their study. Given the prominent use of CSTs in admissions procedures, these contrasting findings suggests a need to study score differences on CSTs between ethnic groups in multiple contexts.

The results of the analyses of predictive validity and ethnic-based score differences highlighted a classic diversity-validity dilemma for the admissions instruments prior academic achievement and the assessments of motivation (Pyburn et al., 2008). Prior academic achievement is a valid predictor of academic success in terms of course grades but substantially lowers ethnic minorities admissions scores. Adding the motivation assessments to the final composite score reduced ethnic-based score differences but substantially decreased the predictive validity of the selection procedure. The results of the CST analyses indicate that this instrument may reduce ethnic-based score difference, while maintaining predictive validity. As such, the CST might offer an alternative to pu-GPA and noncognitive admissions instruments, such as assessments of motivation, as it had smaller negative effects on student population diversity than pu-GPA, and showed better predictive validity than the assessments of motivation.

The result that motivation assessments are poor predictors of academic achievement is common in HE selection research (Niessen & Meijer, 2017; Woo et al., 2023). In particular, the study motivation questionnaire, one of the motivation assessments, performed poorly regarding predictive validity. A plausible explanation for this finding is that some applicants faked their answers, and their answers thus did not reflect their actual motivation to study in HE (Peterson et al., 2011). This may be especially true for applicants with relatively low pu-GPAs, who may attempt to compensate by presenting a high motivation. If questionnaires are used in high-stakes selection contexts, a possible solution to the problem of socially desirable responses might be to use questionnaires with a forced-choice design. Another possible explanation is the low internal consistency of the motivation questionnaire's scales, which is particularly concerning for a high-stakes selection procedure. Assessing motivation or other noncognitive indicators in a high-stakes selection context thus requires higher psychometric standards. Overall, our results restate the importance of researching admissions instruments in real, high-stakes selection contexts.

Regarding differential prediction, our results regarding the overprediction of ethnic minority students' course grades by pre-university grades is in line with prior research on differential

prediction of grade-based selection (Kennet-Cohen et al., 2014; Shewach et al., 2017). Less research has been conducted on differential prediction of motivational instruments and CSTs, especially in high-stakes contexts. Nevertheless, our results suggest that, if differential prediction occurs on these instruments, it occurs in a similar direction as the observed differential prediction of preuniversity grades, namely overprediction of ethnic minority students' course grades. Regarding the CST, prior research has shown that this instrument reduces gender-based differential prediction, in the sense of underprediction of female performance, arguably because the CST is proximal to actual performance in a HE course programme (Niessen et al., 2019). Important here, however, is that none of the instruments in our study showed underprediction of ethnic minority students' course grades. This is considered particularly problematic for HEIs' goals to widening access to HE, because regression lines would disadvantage the group that already has lower access to HE if they underpredict ethnic minority students' grades (Kennet-Cohen et al., 2014).

A limitation of our study was the omission of ethnic background data of applicants who had not enrolled in the study programme, because those data were only available for enrolled students in the university's education research database. Our analyses of ethnic-based score differences therefore relied on data from a sample of selected students. The negative consequences of calculating subscore differences within the sample of enrolled students was partly mitigated, because many applicants who were eventually admitted in the programme, based on their assessment score, declined their invitation to enrol, with the lowest ranking number that was invited to enrol in one of the 600 slots being 975. This increased the range of admission-scores that could be linked to study outcomes. For future research, a recommendation is to use background data of the whole applicant sample, if available (Fikrat-Wevers et al., 2023; Stegers-Jager et al., 2015).

Another way in which we attempted to work around the data omission was by comparing ethnic diversity of the cohort enrolled with selective admissions with two prior cohorts enrolled without selective admissions. These analyses indicated decreased ethnic diversity in the selective admissions cohort. It should be noted however, that this decrease can only partly be explained by ethnic-based score differences on the admissions instruments. Prior research indicated that, compared with ethnic majority students, ethnic minority students apply for selective study programmes less often, because they estimate that their chances of success will be lower (i.e., self-selection; Freeman et al., 2016; Mulder et al., 2022). To overcome this inequality of opportunity in HE due to selective admissions and to increase student population diversity, lottery-based selection has been suggested (Mulder et al., 2022).

A second limitation of our study is the low level of reliability of the CST, which can mostly be explained by (1) the heterogeneous content of the test, in which questions were included on three subfields of psychology and statistics, (2) the relatively low number of items included in the test. Although we believe that the content of the CST was of high quality, because of its alignment with the content and learning objectives of the first year of the study programme and its positive association with Year-1 course grades, practitioners and researchers involved in developing CSTs should strive for more reliable tests, for instance by making the test longer.

A third limitation of our study is the relatively small size of our sample, particularly for detecting slope differences with small effect sizes in the differential prediction analyses. Although our decision to use data that were collected in an actual high-stakes selection context increased the external validity of our results, this also meant that we had to work with a sample whose size was determined based on the number of students that participated in the selection procedure and had enrolled into the course programme. Further research with large sample sizes and data collected in high-stakes contexts is therefore needed to confirm our findings.

A fourth limitation is that the generalisability of our findings may be limited because the 2020-sample had studied their first year in HE during the COVID-19 crisis, leading to drastic changes in both campus life and examination policies, which may have impacted students' academic achievement and the associated inequities between ethnic groups (Goudeau et al., 2021). Future research should investigate ethnic-based score differences and predictive validity of selective admissions in other study programmes and not under COVID restricted conditions. Future research should also include gender, socioeconomic background, or intersections between various social categories. Lastly, future research should investigate performance on the various admissions instruments of students who are the first in their family to study in HE (i.e., first-generation in HE students).

We conclude that admitting HE students based on a composite score which includes an assessment of motivation alongside prior academic achievement may mitigate ethnic-based score differences. However, these motivation assessments are poor predictors of academic success, especially compared to pu-GPA. Our study further demonstrates that CSTs, as compared to pu-GPA, show promise in lowering ethnic-based score differences as compared to pu-GPA, while maintaining substantial predictive validity for both ethnic majority and ethnic minority students.

AUTHOR CONTRIBUTIONS

Pieter van Lamoen is a PhD student at the Department of Psychology, Education & Child Studies at Erasmus University Rotterdam, the Netherlands. His research interests include the transition from secondary to higher education, student diversity, and academic success. Annemarie Hiemstra is an Associate Professor at the Department of Psychology, Education and Child Studies of the Erasmus University Rotterdam, the Netherlands. Her research interests include personnel selection, psychological assessment, hiring discrimination and workplace diversity. Marieke Meeuwisse is an Associate Professor at the Department of Psychology, Education and Child Studies of the Erasmus University Rotterdam, the Netherlands. Central themes in her research are diversity, inclusion, and student success, with a special interest in the higher education learning environment, student and staff interactions and sense of belonging. Lidia Arends is Professor of Methodology and Statistics at the Department of

Psychology, Education & Child Studies, Erasmus University Rotterdam, the Netherlands. Besides, she is biostatistician at the Department of Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands. Her areas of interest include research methods, (logistic) regression analysis, multilevel analysis, systematic reviews, and meta-analysis. Sabine Severiens is a Full Professor at the Department of Psychology, Education and Child Studies of the Erasmus University Rotterdam, the Netherlands. Her research interests include diversity and educational equity, from the perspective of the teacher (culturally responsive teaching) as well as the student (motivation, integration, and the learning environment).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Pieter M. van Lamoen http://orcid.org/0000-0003-0209-8315

REFERENCES

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, *95*, 648–680. https://doi.org/10.1037/a0018714

Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 435–463. https://doi.org/10.1146/annurev-orgpsych-032414-111256

de Boer, T., & van Rijnsoever, F. (2022). In search of valid non-cognitive student selection criteria. *Assessment & Evaluation in Higher Education*, *47*(5), 783–800. https://doi.org/10.1080/02602938.2021.1958142

Bowman, N. A. (2010). College diversity experiences and cognitive development: A meta-analysis. *Review of Educational Research*, *80*(1), 4–33. https://doi.org/10.3102/0034654309352495

Bradburn, J., & Schmitt, N. (2019). Combining cognitive and noncognitive predictors and impact on selected individual demographics: An illustration. *International Journal of Selection and Assessment*, *27*(1), 21–30. https://doi.org/10.1111/ijsa.12234

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, *5*(2), 115–124. https://doi.org/10.1111/j.1745-3984.1968.tb00613.x

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*(6), 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

Dahlke, J. A., & Sackett, P. R. (2018). Refinements to effect sizes for tests of categorical moderation and differential prediction. *Organizational Research Methods*, *21*(1), 226–234. https://doi.org/10.1177/1094428117736591

Dahlke, J. A., Sackett, P. R., & Kuncel, N. R. (2019). Effects of range restriction and criterion contamination on differential validity of the SAT by race/ethnicity and sex. *Journal of Applied Psychology*, *104*(6), 814–831. https://doi.org/10.1037/apl0000382

Dutch Inspectorate of Education. (2023). *Selectie in het hoger onderwijs: Criteria, instrumenten en de borging van kansengelijkheid [Selection in higher education: Criteria, tools and ensuring equity of opportunity]*. Dutch Inspectorate of Education. https://www.onderwijsinspectie.nl/documenten/rapporten/2023/01/20/selectie-in-het-hoger-onderwijs-criteria-instrumenten-en-de-borging-van-kansengelijkheid

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642. https://doi.org/10.1007/s00038-012-0416-3

Fikrat-Wevers, S., Stegers-Jager, K. M., Afonso, P. M., Koster, A. S., van Gestel, R. A., Groenier, M., Ravesloot, J. H., Wouters, A., Van Den Broek, W. W., & Woltman, A. M. (2023). Selection tools and student diversity in health professions education: A multi-site study. *Advances in Health Sciences Education*, *28*, 1027–1052. https://doi.org/10.1007/s10459-022-10204-9

Freeman, B. K., Landry, A., Trevino, R., Grande, D., & Shea, J. A. (2016). Understanding the leaky pipeline: Perceived barriers to pursuing a career in medicine or dentistry among underrepresented-in-medicine undergraduate students. *Academic Medicine*, *91*(7), 987–993.

Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour*, *5*(10), 1273–1281. https://doi.org/10.1038/s41562-021-01212-7

Gurin, P., Dey, E., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, *72*(3), 330–367. https://doi.org/10.17763/haer.72.3.01151786u134n051

Hill, N. E., & Wang, M. T. (2015). From middle school to college: Developing aspirations, promoting engagement, and indirect pathways from parenting to post high school enrollment. *Developmental Psychology*, *51*(2), 224–235. https://doi.org/10.1037/a0038367

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*(1–2), 152–194. https://doi.org/10.1111/1468-2389.00171

Isik, U., Tahir, O. E., Meeter, M., Heymans, M. W., Jansma, E. P., Croiset, G., & Kusurkar, R. A. (2018). Factors influencing academic motivation of ethnic minority students: A review. *SAGE Open*, *8*(2), 215824401878541. https://doi.org/10.1177/2158244018785412

Isik, U., Wouters, A., ter Wee, M. M., Croiset, G., & Kusurkar, R. A. (2017). Motivation and academic performance of medical students from ethnic minorities and majority: A comparative study. *BMC Medical Education*, *17*(1), 233. https://doi.org/10.1186/s12909-017-1079-9

Kennet-Cohen, T., Turvall, E., & Oren, C. (2014). Detecting bias in selection for higher education: Three different methods. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 193–204. https://doi.org/10.1080/0969594X.2013.877871

Kickert, R., Meeuwisse, M., Arends, L. R., Prinzie, P., & Stegers-Jager, K. M. (2021). Assessment policies and academic progress: Differences in performance and selection for progress. *Assessment & Evaluation in Higher Education*, *46*(7), 1140–1156. https://doi.org/10.1080/02602938.2020.1845607

Kurysheva, A., van Rijen, H. V. M., & Dilaver, G. (2019). How do admission committees select? Do applicants know how they select? Selection criteria and transparency at a Dutch University. *Tertiary education and management*, *25*(4), 367–388. https://doi.org/10.1007/s11233-019-09050-z

Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement*, *10*(2), 133–139. https://doi.org/10.1177/014662168601000202

Lawley, D. N. (1944). IV.—A note on Karl Pearson's selection formulæ. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, *62*(1), 28–30. https://doi.org/10.1017/S0080454100006385

Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct

validity. *International Journal of Selection and Assessment*, *10*(4), 245–257. https://doi.org/10.1111/1468-2389.00215

Mulder, L., Akwiwu, E. U., Twisk, J. W. R., Koster, A. S., Ravesloot, J. H., Croiset, G., Kusurkar, R. A., & Wouters, A. (2023). Inequality of opportunity in selection procedures limits diversity in higher education: An intersectional study of Dutch selective higher education programs. *PLoS One*, *18*(10), e0292805. https://doi.org/10.1371/journal.pone.0292805

Mulder, L., Wouters, A., Twisk, J. W. R., Koster, A. S., Akwiwu, E. U., Ravesloot, J. H., Croiset, G., & Kusurkar, R. A. (2022). Selection for health professions education leads to increased inequality of opportunity and decreased student diversity in The Netherlands, but lottery is no solution: A retrospective multi-cohort study. *Medical Teacher*, *44*(7), 790–799. https://doi.org/10.1080/0142159x.2022.2041189

Niessen, A. S. M., & Meijer, R. R. (2017). On the use of broadened admission criteria in higher education. *Perspectives on Psychological Science*, *12*(3), 436–448. https://doi.org/10.1177/1745691616683050

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Predicting performance in higher education using proximal predictors. *PLoS One*, *11*(4), e0153663. https://doi.org/10.1371/journal.pone.0153663

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2019). Gender-based differential prediction by curriculum samples for college admissions. *Educational Measurement: Issues and Practice*, *38*(3), 33–45. https://doi.org/10.1111/emip.12266

Nieuwhof, M. G. H., ThJ ten Cate, O., Oosterveld, P., & Soethout, M. B. M. (2004). Measuring strength of motivation for medical school. *Medical Education Online*, *9*(1), 4355. https://doi.org/10.3402/meo.v9i.4355

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Nye, C. D., & Sackett, P. R. (2017). New effect sizes for tests of categorical moderation and differential prediction. *Organizational Research Methods*, *20*(4), 639–664. https://doi.org/10.1177/1094428116644505

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, *24*(3), 270–290. https://doi.org/10.1080/08959285.2011.580808

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racio-ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*(1), 153–172. https://doi.org/10.1111/j.1744-6570.2008.00109.x

Pyburn, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity–validity dilemma: Overview and legal context. *Personnel Psychology*, *61*(1), 143–151. https://doi.org/10.1111/j.1744-6570.2008.00108.x

Reibnegger, G., Caluba, H.-C., Ithaler, D., Manhal, S., Neges, H. M., & Smolle, J. (2010). Progress of medical students after open admission or admission based on knowledge tests. *Medical Education*, *44*(2), 205–214. https://doi.org/10.1111/j.1365-2923.2009.03576.x

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rosinger, K. O., Sarita Ford, K., & Choi, J. (2021). The role of selective college admissions criteria in interrupting or reproducing racial and economic inequities. *The Journal of Higher Education*, *92*(1), 31–55. https://doi.org/10.1080/00221546.2020.1795504

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*(6), 1479–1497.

Shewach, O. R., Shen, W., Sackett, P. R., & Kuncel, N. R. (2017). Differential prediction in the use of the SAT and high school grades in predicting college performance: Joint effects of race and language. *Educational Measurement: Issues and Practice*, *36*(3), 46–57. https://doi.org/10.1111/emip.12150

Soppe, K. F. B., Klugkist, I. G., Wubbels, T., & de Meij, L. D. N. V. W. (2024). Determining fit: The role of matching procedures in prospective higher education students' enrolment behaviour. *Journal of Higher Education Policy and Management*, *46*(2), 131–145. https://doi.org/10.1080/1360080X.2023.2267201

Stegers-Jager, K. M. (2018). Lessons learned from 15 years of non-grades-based selection for medical school. *Medical Education*, *52*(1), 86–95. https://doi.org/10.1111/medu.13462

Stegers-Jager, K. M., Steyerberg, E. W., Lucieer, S. M., & Themmen, A. P. N. (2015). Ethnic and social disparities in performance on medical school selection criteria. *Medical Education*, *49*(1), 124–133. https://doi.org/10.1111/medu.12536

Thorndike, R. L. (1949). *Personnel selection; test and measurement techniques* (Vol. 358, p. viii). Wiley.

de Visser, M., Fluit, C., Cohen-Schotanus, J., & Laan, R. (2018). The effects of a non-cognitive versus cognitive admission procedure within cohorts in one medical school. *Advances in Health Sciences Education*, *23*(1), 187–200. https://doi.org/10.1007/s10459-017-9782-1

Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education*, *43*(3), 399–414. https://doi.org/10.1080/02602938.2017.1353586

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*(5), 372–376. https://doi.org/10.1037/h0026244

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, *3*(1), 94–123. https://doi.org/10.1177/2515245919885611

Woo, S. E., LeBreton, J. M., Keith, M. G., & Tay, L. (2023). Bias, fairness, and validity in graduate-school admissions: A psychometric perspective. *Perspectives on Psychological Science*, *18*(1), 3–31. https://doi.org/10.1177/17456916211055374

Wouters, A., Croiset, G., Galindo-Garre, F., & Kusurkar, R. A. (2016). Motivation of medical students: Selection by motivation or motivation by selection. *BMC Medical Education*, *16*(1), 37. https://doi.org/10.1186/s12909-016-0560-1

Zamanzadeh, V., Ghahramanian, A., Valizadeh, L., Bagheriyeh, F., & Lynagh, M. (2020). A scoping review of admission criteria and selection methods in nursing education. *BMC Nursing*, *19*, 121. https://doi.org/10.1186/s12912-020-00510-1

Zimmermann, S., Klusmann, D., & Hampe, W. (2017). Correcting the predictive validity of a selection test for the effect of indirect range restriction. *BMC Medical Education*, *17*(1), 246. https://doi.org/10.1186/s12909-017-1070-5

Zwick, R. (2019). Fairness in measurement and selection: Statistical, philosophical, and public perspectives. *Educational Measurement: Issues and Practice*, *38*(4), 34–41. https://doi.org/10.1111/emip.12299

How to cite this article: Lamoen, P. M., Hiemstra, A. M. F., Meeuwisse, M., Arends, L. R., & Severiens, S. E. (2024). Ethnic-based score differences and differential prediction of various cognitive and noncognitive admissions instruments in higher education. *International Journal of Selection and Assessment*, 1–14. https://doi.org/10.1111/ijsa.12501