Personality and Individual Differences 123 (2018) 50–55

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Differentiation of cognitive abilities and the Medical College Admission Test

Matthew J.W. McLarnon[a,*], Richard D. Goffin[b], Mitchell G. Rothstein[c]

[a] Department of Psychology, Oakland University, United States
[b] Department of Psychology, University of Western Ontario, Canada
[c] Department of Management and Organizational Studies, University of Western Ontario, Canada

ARTICLE INFO

Keywords:
Cognitive ability
Ability differentiation
Medical College Admission Test (MCAT)
Measurement invariance

ABSTRACT

Cognitive Ability Differentiation (CAD) theory proposes greater differentiation of narrow cognitive abilities occurs at a high level of *g*, the general factor of ability. CAD also proposes that *g* is a stronger determinant of cognitive performance in low-*g* versus high-*g* individuals, but that narrow cognitive abilities are stronger determinants of performance in high- versus low-*g* individuals. We assessed whether CAD occurs in Medical College Admission Test (MCAT) scores using data from over 4800 medical school applicants. In support of CAD, our results provided consistent evidence that the MCAT's measures of narrow cognitive abilities were more differentiated in high- versus low-*g* individuals. Also consistent with CAD, *g* was a stronger predictor of a cognitive performance criterion, GPA, in low- versus high-*g* individuals. Contrary to CAD, however, the MCAT's measures of narrow abilities were not stronger predictors of GPA in high- versus low-*g* individuals. Implications of these results for future CAD research and for medical schools' use of the MCAT are discussed.

Spearman (1927) observed that correlations between tests of different cognitive abilities tended to be weaker in individuals with higher intelligence. A theory of cognitive ability differentiation (CAD) has been developed from this premise (Jensen, 2003). Many investigations into CAD have been conducted, but have resulted in mixed findings. The two main tenets of CAD are 1) greater differentiation among narrow cognitive abilities at a high versus low level of *g* (Abad, Colom, Juan-Espinosa, & García, 2003), and 2) stronger predictiveness of cognitive performance by *g* in low- versus high-*g* individuals, but stronger predictiveness of cognitive performance by narrow cognitive abilities in high- versus low-*g* individuals (Murray, Dixon, & Johnson, 2013). In a unique addition to previous investigations, we assessed whether CAD occurs in applicants' scores on the Medical College Admission Test (MCAT).

Tenet 1 of CAD suggests differences in *g*-factor loadings and residual variances with regards to the narrow cognitive abilities, and overall *g*-factor variances across high- and low-*g* groups. Loadings of narrow cognitive abilities on *g* (λs in Fig. 1) should be lower in a group of applicants with high versus low *g*. This signifies that at higher *g*, *g* contributes less to narrow cognitive abilities, and that relations between narrow abilities are weaker at higher levels of *g* (Jensen, 2003). Correspondingly, CAD proposes that the variance in narrow abilities that is not explained by *g* (residual variance) will be greater in the high- versus low-*g* group as the narrow abilities become more varied in individuals with greater *g*. Additionally, given that less variability is associated with narrow cognitive abilities at low *g*, the variance associated with the *g*-factor should be greater in these individuals. Further, *g* should explain less variance in MCAT scores in high-*g* individuals.

Tenet 2 suggests that the relation between *g* and cognitive performance criteria will be stronger in the low- versus high-*g* group. Conversely, Tenet 2 also suggests that the relations between narrow abilities and cognitive performance criteria will be stronger in the high- versus low-*g* group. To assess Tenet 2, we assessed differences across low- and high-*g* groups in the relations of *g*, and the narrow abilities assessed by the MCAT, with medical school applicants' grade point average (GPA).

In the next section, we review a number of studies that have examined CAD. Almost all previous studies have been conducted in low-stakes testing situations, but it is well-known that high-stakes testing situations, such as our MCAT testing scenario, can substantially change fundamental structural and validational aspects of test scores (McLarnon, Goffin, Schneider, & Johnston, 2016). Also, interestingly, few studies have examined Tenet 2.

1. Research supporting cognitive ability differentiation

The following research findings stem from both individual- and group-level studies, suggesting that CAD may be observed at multiple levels. As examples of individual-level findings, Detterman and Daniel (1989) compared the correlations between narrow cognitive ability

* Corresponding author.
E-mail address: mclarnon@oakland.edu (M.J.W. McLarnon).

https://doi.org/10.1016/j.paid.2017.11.005
Received 15 June 2017; Received in revised form 3 November 2017; Accepted 6 November 2017

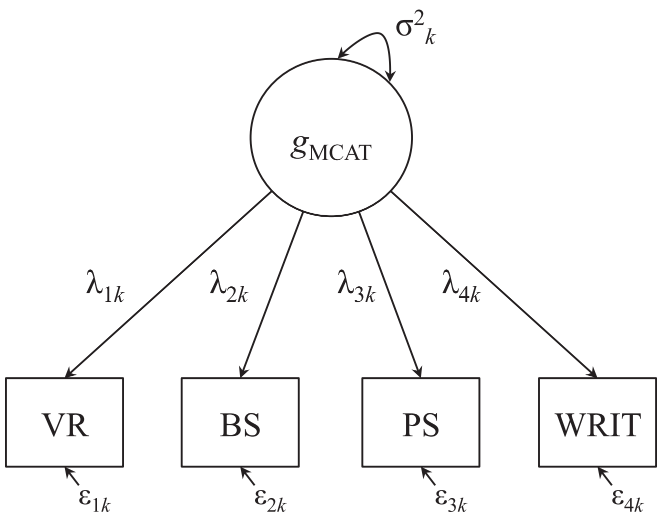

Fig. 1. Conceptual *g*-factor model defined by subtests of Medical College Admissions Test (MCAT). VR = verbal reasoning, BS = biological science, PS = physical science, WRIT = writing. Parameters that were tested for invariance across low- and high-*g* groups are given by *k* subscripts.

measures across low- and high-*g* students, and found that in the low-*g* group, correlations were approximately 0.50, whereas in the high-*g* group, correlations were about 0.30 (see also Legree, Pifer, & Grafton, 1996). Using principal component analysis, Deary et al. (1996) found that *g* explained substantially more variance in narrow abilities at low *g*. These findings demonstrate consistency across different samples and cognitive measures. In terms of the group-level findings, Coyle and Rindermann (2013) found that low-*g* nations had substantially lower *g*-factor loadings for narrow abilities compared to high-*g* nations. Similarly, Nyborg and Jensen (2000) observed that differentiation was independent from test bias across low- and high-*g* Caucasian and African-American test-takers.

Additionally, studies using unique individual-level samples have also provided evidence for CAD. For example, Blanch, García, Llaveria, and Aluja (2017) found unequal *g*-factor loadings and residual variances across amateur and expert chess players, thereby supporting the presence of differentiation.

2. Research failing to support cognitive ability differentiation

Potentially due to a series of limitations highlighted in this section, other studies have failed to support CAD. Fogarty and Stankov (1995) suggested that supportive findings were associated with high-*g* individuals finding the tasks associated with ability measures too easy, reducing correlations between narrow abilities. Hartmann and Reuter (2006) critiqued the practice of dichotomizing *g* because it can result in arbitrary separations and loss of power. However, McLarnon and Carswell (2013) suggested that a discrete-group approach might be one of the few viable options to study CAD. In particular, using an extreme-group approach where the top- and bottom-thirds are separated into discrete groups can help address the issue of arbitrary divides between individuals in adjacent groups (Kline, 2005). Results of Saklofske, Yang, Zhu, and Austin (2008) did not support CAD, however, their study only compared adjacent ability groups, potentially leading to an arbitrary distinction between high and low-*g* individuals. Further, with sufficiently large samples the reduction in power associated with extreme groups may be minimized. Notably, small sample size may have impacted Fogarty and Stankov's (1995) suggestion of ceiling effects, given that their high- and low-*g* groups consisted of $n < 30$. Furthermore, the MCAT would likely be seen as difficult for test-takers of all ability levels. Thus, to help address limitations of previous CAD research, we used an extreme-group approach in which the size of each group was substantial, and the test was sufficiently difficult.

In light of concerns over dichotomizing *g*, other studies have attempted to leverage new data analysis techniques to investigate CAD. Murray et al. (2013) used moderated factor analysis to explore whether the level of the latent factor moderates factor loadings. Although Murray et al.'s findings did not support CAD, further exploration in unique testing situations may be necessary. Furthermore, a potential limitation of moderated factor models is that it is not currently possible to estimate variability of residual variances (see Bauer, 2016). Thus, in the current study we use an extreme group approach to provide a thorough assessment of CAD's tenets. In Online Supplemental Material we describe the results of the moderated factor model, which demonstrate consistency with the results presented below.

Reynolds, Keith, and Beretvas (2010) examined differentiation by applying factor mixture analysis (FMA). FMA is a type of mixture model, which refers to the concept that data may represent a 'mix' of parameters (e.g., means, factor loadings). Reynolds et al.'s results partially supported Tenet 1, in that *g*-variance was lower at high-*g*. However, their study did not investigate whether relations between narrow abilities were weaker at high- versus low-*g*. This may be inherent with FMA, as typical applications constrain factor loadings to equality across latent classes (McLarnon, Carswell, & Schneider, 2015). Although it may be possible to specify a FMA with unequal factor loadings across classes, such a model may often be empirically under-identified, likely requiring extremely large sample sizes. Thus, FMA may not be an ideal statistical method for differentiation investigations.

3. Current study

Using multi-group confirmatory factor analysis (CFA) we investigated CAD. This approach presents a straightforward methodology that leverages measurement invariance analyses to sequentially test the equivalence of parameters associated with CAD (Carlstedt, 2001; Marsh et al., 2010; McLarnon & Carswell, 2013). This study makes four contributions. First, we contribute a unique examination of CAD. Specifically, we included the data of every candidate who completed the MCAT when applying to medical school programs in a large province in Canada during the years of 2004, 2006, 2007, and 2008. This provided a total sample size exceeding 7000, from which we derived a low-*g* (bottom third) and a high-*g* (top third) group.

Second, we followed Marsh et al.'s (2010) recommendations (see *Analytical procedure*) for examining measurement invariance across groups. This facilitated a straightforward examination of the equality of factor loadings, residual variances, and factor variances across the *g*-groups to assess Tenet 1.

Third, we provide an investigation of differentiation using high-stakes testing data. High-stakes situations are ones in which a test is completed for the purpose of obtaining a highly-desired outcome (i.e., gaining admittance to medical school). Few, if any, previous investigations have examined whether CAD may occur in situations involving high-stakes testing.

Fourth, to assess Tenet 2, we examined relations of the MCAT's *g*-factor and narrow abilities scales with medical school applicants' GPA. As GPA is also one of the focal indicators of success in universities it serves as a suitable cognitive performance criterion in the assessment of predictive hypotheses[1] (McLarnon et al., 2017).

In sum, we investigated two general hypotheses:

Hypothesis 1. Tests of equivalence of factor loadings, residual variances, and factor variances across low- and high-*g* groups of medical school applicants will reveal a lack of invariance, thereby supporting Tenet 1.

[1] We use the terms *predictor* and *predictiveness* for descriptive purposes, and do not imply causality.

Table 1
Measurement invariance analyses.

Model	χ^2	χ^2c	χ^2 *df*	CFI	RMSEA (90% CI)	Model comparison	$\Delta\chi^2$	$\Delta\chi^2$ *df*	ΔCFI	ΔRMSEA
1	57.35*	0.88	4	0.954	0.074 (0.058–0.092)	–	–	–	–	–
2	93.33*	1.04	7	0.926	0.071 (0.059–0.085)	2 vs. 1	37.18*	3	− 0.028	− 0.003
3	498.75*	0.96	11	0.583	0.135 (0.125–0.146)	3 vs. 2	467.89*	4	− 0.343	0.064
						3 vs. 1	426.30*	7	− 0.371	0.061
3b	499.13*	0.88	8	0.580	0.159 (0.147–0.171)	3b vs. 1	442.71*	4	− 0.374	0.085
4	176.58*	1.06	8	0.856	0.093 (0.082–0.105)	4 vs. 2	77.58*	1	− 0.070	0.022
						4 vs. 1	110.48*	4	− 0.098	0.019
4b	72.76*	0.92	5	0.942	0.075 (0.060–0.090)	4b vs. 1	15.26*	1	− 0.012	0.001
5	605.74*	1.00	12	0.492	0.143 (0.133–0.153)	5 vs. 4	471.13*	4	− 0.364	0.050
						5 vs. 3	87.82*	1	− 0.091	0.008
						5 vs. 2	539.93*	5	− 0.434	0.072
						5 vs. 1	524.08*	8	− 0.462	0.069

Note. n_{low} = 2421 n_{high} = 2425. χ^2c = χ^2 scaling correction; *df* = degrees of freedom; #fp = number of parameters estimated; CFI = comparative fit index; RMSEA = root mean square error of approximation; $\Delta\chi^2$ = scaled χ^2 difference statistic; ΔCFI = change in CFI estimate from less restricted to more restricted models; ΔRMSEA = change in RMSEA estimate. Model 1 = configural invariance, Model 2 = metric/factor loading invariance, Model 3 = metric + uniqueness invariance, Model 3b = uniqueness invariance, Model 4 = metric + factor variance invariance, Model 4b = factor variance invariance, Model 5 = metric + uniqueness + factor variance invariance.
* $p < 0.01$.

Hypothesis 2a. The *g*-factor will demonstrate stronger predictiveness of cognitive performance (GPA) in the low- versus the high-*g* group.

Hypothesis 2b. The MCAT's narrow abilities will demonstrate stronger predictiveness of cognitive performance (GPA) in the high- versus the low-*g* group.

4. Method

4.1. Participants

Participants were all 7498 applicants to all of the medical schools in one large Canadian province during the years of 2004 and 2006–2008. Demographics were not available for all participants, but in a subsample (n = 330) used in McLarnon et al. (2017), 59.1% were female, and were on average 25.00 years old (*SD* = 2.05). It is likely that the demographics of that subsample are representative of the larger sample.

Low-*g* (n = 2421) and high-*g* (n = 2425) groups were based on the lower and upper tertiles defined by the overall MCAT score. The middle third was omitted from further analysis as the use of the extreme-group approach was deemed the most appropriate option for this study.

4.2. Measures

The MCAT (AAMC, 1991) consists of four subtests: verbal reasoning, physical sciences, biological sciences, and writing. The MCAT's validity has been established in a wide-range of previous research (Julian, 2005; McGaghie, 2002).

As is typical, the average GPA of the applicants' best two years of undergraduate study served as a measure of GPA (McGaghie, 2002), and was used as the cognitive performance criterion in this study.

4.3. Analytical procedure

Our focal model was a first-order CFA, with a single latent variable representing the *g*-factor (Fig. 1). All four of the MCAT subtests were treated as indicators of the latent *g*-factor. We followed Marsh et al.'s (2010) invariance testing procedure, which required that we test and compare the fit of models that varied in their invariance assumptions. In Model 1, configural invariance was specified in order to assess whether the same factor model, with no equality constraints on factor loadings, residual variances, or factor variances, was supported across groups.

Model 2 specified metric invariance by assessing whether respective MCAT abilities have equal factor loadings across groups. Model 3 added equality constraints on respective residual variances. Model 4 was tested by placing an equality constraint on the variance of the latent factor across groups, over and above metric invariance, without residual variance constraints. Model 5 specified a complete invariance model: corresponding factor loadings, residual variances, and *g*-factor variance were constrained to equality across groups. We also assessed residual variance equality (Model 3b) and factor variance equality (Model 4b) without the metric invariance constraints to investigate whether those constraints were individually equivalent across *g*-groups (Marsh et al., 2010). We did not consider tests of intercept equality because unequal intercepts were confounded with the *g*-based groups.

Fit was assessed using: comparative fit index (CFI) values of > 0.90 and > 0.95, and root mean square error of approximation (RMSEA) values of < 0.08 and < 0.05, to indicate adequate and good fit respectively (Goffin, 2007; Hu & Bentler, 1999). Overall model fit estimates were supplemented with maximal reliability (H) to provide an indication of measurement quality (McNeish, An, & Hancock, in press). Invariance is supported by ΔCFI ≤ 0.010 and/or ΔRMSEA ≤ 0.015 versus a preceding model (Sass, 2011). The equality of the predictiveness of GPA across groups, Wald χ^2 tests from M*plus* 7.4's (Muthén & Muthén, 2012) MODEL TEST procedure were examined. M*plus*' robust maximum likelihood estimator was used.

5. Results

Online Supplemental Material contains descriptives and correlations for the full sample and each *g*-group. Preliminary analyses using the full sample suggested adequate measurement quality of the MCAT factor by its four narrow ability measures, H = 0.83, $p < 0.01$ (McNeish et al., in press).

5.1. Tenet 1: assessing measurement invariance

Table 1 presents model fit indices for the invariance analyses. Model 1, configural invariance, demonstrated adequate fit (CFI = 0.95, RMSEA = 0.07). Table 2 provides standardized estimates from the configural model for high- and low-*g* groups. Adding metric invariance constraints, the fit of Model 2 was substantially worse according to ΔCFI (Table 1). Failing metric invariance suggested that relations between narrow abilities and the *g*-factor were discrepant across groups. Factor loadings for verbal reasoning, biological science, and physical science were significantly stronger in the low- versus high-*g* group. However, contrary to expectations, the narrow ability of writing had a significantly stronger loading in the high- versus low-*g* group. Nonetheless, metric invariance was not supported because of stronger

Table 2
Configural invariance (baseline) model estimates.

	Low				High			
	λ	*SE*	ε	*SE*	λ	*SE*	ε	*SE*
Verbal reasoning	0.20*	0.03	0.96*	0.01	0.05	0.03	1.00*	0.01
Physical science	0.64*	0.04	0.59*	0.06	0.46*	0.03	0.79*	0.03
Writing	− 0.26*	0.04	0.93*	0.02	− 0.45*	0.04	0.80*	0.03
Biological science	0.76*	0.05	0.43*	0.08	0.57*	0.04	0.67*	0.04
	σ^2	*SE*			σ^2	*SE*		
g	0.07*	0.01			0.01*	0.01		

Note. The top panel gives standardized factor loadings, λ, and residual variances, ε, and their respective standard errors (SEs) from the configural invariance model. The bottom panel gives the variance, σ^2, of the latent *g*-factor from the metric invariance model.
* $p < 0.01$.

relations among MCAT subtests in the low- versus high-*g* group, which supports Tenet 1.

The poor overall fit of Model 3, and its ΔCFI and ΔRMSEA values, suggested a lack of equality of residual variances of the narrow abilities across groups. Model 3b, which constrained the residual variances but freed the factor loadings to vary across groups, also fit poorly. These results suggested greater variability in narrow abilities at higher *g*, supporting Tenet 1.

Model 4, which constrained factor loadings as well as factor variances to equality across groups, was not supported by the ΔCFI and ΔRMSEA criteria. Model 4b, which omitted metric invariance but still constrained factor variances, was also not supported. In both cases, the unacceptable ΔCFI and ΔRMSEA were due to substantially larger *g*-factor variance in the low- versus high-*g* group. This further supported Tenet 1.

Finally, the constraints associated with equal factor loadings, residual variances, and factor variances were combined in Model 5. Model 5′s poor fit further supported the lack of invariance across groups, which supports Tenet 1. Together, these analyses supported the plausibility of Hypothesis 1.[2]

5.2. Tenet 2: assessing relations with GPA

Tenet 2 proposed that the relation between *g* and the cognitive performance criterion, GPA, will be stronger in the low- versus high-*g* group, and that the relations between narrow abilities and cognitive performance will be stronger in the high- versus low-*g* group. To test this we compared the correlations of *g* and the MCAT's narrow abilities with GPA in a sequential manner (i.e., each of the following tests were derived in separate analyses). Rather than using raw correlations for this assessment, results were drawn from the Model 1 because of its superior overall fit. Results are presented in Table 3.[3] *g* and GPA correlated at 0.51, $p < 0.01$ in the low-*g* group, and 0.36, $p < 0.01$ in the high-*g* group. A significant Wald test emphasized that the association between *g* and GPA was significantly stronger in the low- versus high-*g* group. This supported Tenet 2 and Hypothesis 2a.

To further assess Tenet 2, we tested the hypothesis that the correlation between each of the narrow abilities and GPA would be stronger in the high- versus low-*g* group (see Table 3). The correlations for verbal reasoning were found to be 0.03, $p = 0.17$, and − 0.02, $p = 0.36$, in the low- and high-*g* groups, respectively, and were not significantly different from each other. For physical science, the correlation with GPA was significantly stronger (0.31, $p < 0.01$) in the low- compared

[2] Online Supplemental Material presents an application of the moderated factor model, which demonstrate consistency with the focal results.

[3] There was moderate negative skew in GPA, in both groups, resulting in censored data. Analyses were also conducted using M*plus*' censored variable and tobit regression procedures (Muthén & Muthén, 2012), and log-transforming GPA. Results from these analyses parallel those reported.

Table 3
GPA relations.

	Low	High	$\chi^2(1)$
g	0.51* (0.02)	0.36* (0.03)	65.82*
Verbal reasoning	0.03 (0.03)	− 0.02 (0.02)	2.53
Physical science	0.31* (0.03)	0.19* (0.02)	31.65*
Writing	0.15* (0.03)	0.04 (0.03)	22.23*
Biological science	0.31* (0.03)	0.20* (0.02)	23.80*

Note. Correlations, with associated standard errors in parentheses presented under each *g*-group label. The last column presents Wald χ^2 tests, with $df = 1$.
* $p < 0.01$.

to the high-*g* group (0.19, $p < 0.01$). The correlation for writing was $r = 0.15, p < 0.01$ in the low-*g* group, and 0.04, $p = 0.14$, in the high-*g* group, which were significantly different from each other. Finally, for the relation involving biological science, the correlation was 0.31, $p < 0.01$ in the low-*g* group, and $r = 0.20$, $p < 0.01$ in the high-*g* group, and also differed significantly. Thus, for the physical science, writing, and biological science narrow abilities, there was evidence for differential effects, however, the direction of these effects was opposite to Tenet 2, contradicting Hypothesis 2b.

The generally weaker criterion relations of narrow abilities with GPA in high-*g* participants might be attributable to lower *g*-loadings and therefore lower redundancy among the narrow abilities in this group. However, this lower redundancy might result in a predictive advantage when the narrow abilities are considered as a set rather than individual predictors. Thus, GPA was then regressed onto the four narrow cognitive abilities.[4] In the low-*g* group $R^2 = 0.21, p < 0.01$, and in the high-*g* group $R^2 = 0.08$, $p < 0.01$. These estimates were significantly different, Wald $\chi^2(1) = 4.41$, $p < 0.05$, suggesting that all four narrow abilities, as simultaneous predictors, accounted for greater variability in GPA in the low- than in the high-*g* group. In sum, regardless of whether one considers the MCAT's narrow abilities individually or as a set, their predictiveness of a cognitive criterion was generally greater in the low-*g* versus high-*g* group, contrary to Tenet 2. However, the prediction of a cognitive criterion by the MCAT's *g*-factor was superior in the low- versus high-*g* group, consistent with Tenet 2. In sum, Hypothesis 2a received support, but Hypothesis 2b did not.

6. Discussion

This study focused on cognitive ability differentiation (CAD) as it is reflected in the MCAT and a large sample of medical school applicants. To assess Tenet 1 of CAD we assessed whether factor loadings, residual variances, and *g*-factor variances are invariant across low- and high-*g* groups of applicants. Tenet 2 was assessed by considering the equivalence of the MCAT's *g*-factor and narrow abilities to predict GPA across the high- and low-*g* groups.

6.1. Evaluation of Tenet 1 of cognitive ability differentiation

6.1.1. Invariance testing of g-factor loadings

We examined whether the factor loadings of each narrow ability on a *g*-factor, varied across the high-*g* and low-*g* groups. Three of the MCAT's four narrow abilities (verbal reasoning, biological science, and physical science) demonstrated higher loadings in the low- versus high-*g* group, partially supporting Tenet 1. However, writing ability's *g*-loading was significantly stronger in the high-*g* group. One potential explanation is that the MCAT's writing subscale may be closer to a general versus narrow evaluation of ability. Composing a narrative requires considerable verbal knowledge, but also narrow content

[4] GPA was standardized separately in each group so that its variance was approximately equal across groups. This was necessary because the Wald test assessed the equivalence of the residual variance of GPA (i.e., 1-R^2) across groups.

knowledge, which might saturate writing with additional *g* variance. Interestingly, the writing subtest demonstrated negative factor loadings in both *g*-groups, which stand in contrast to the full-sample correlations (see Online Supplemental Material). Negative relations, however, might actually be expected because of compensatory relations between the subtests that manifest when *g* is split into high-*g* and low-*g* groups (Jensen, 2003).

6.1.2. Invariance testing of residual variances

We also posited that Tenet 1 of CAD could be exhibited through greater residual variances of the narrow abilities at high-*g*. Invariance testing suggested that residual variances of verbal reasoning, biological science, and physical science were indeed greater in the high-*g* group. However, similar to the earlier invariance findings on factor loadings, the writing subtest was the exception, exhibiting smaller residual variance in the high- versus low-*g* group. Nevertheless, overall, Tenet 1 received partial support because residual variances of three out of four MCAT subscales were larger in the high-*g* group.

6.1.3. Invariance testing of *g*-factor variances

Tenet 1 of CAD also implied that *g* should have greater variance in the low-*g* group, and our analyses supported this: the variance of the *g*-factor was significantly greater in the low-*g* group.

Collectively, the preponderance of evidence rendered by the current study supported Tenet 1 of CAD. Thus, with respect to the MCAT, there is, in general, greater differentiation among narrow cognitive abilities within a sample of high- versus low-*g* individuals.

6.2. Evaluation of Tenet 2 of cognitive ability differentiation: GPA relations

Tenet 2 suggested that the MCAT's *g*-factor should be more predictive of cognitive performance in the low- versus high-*g* group, and that narrow abilities should be stronger predictors in the high- versus low-*g* group. A significantly stronger correlation between *g* and GPA was found in the low- versus high-*g* group. However, relations of the narrow MCAT abilities with GPA were also generally stronger in the low- versus high-*g* group. This occurred regardless of whether the narrow cognitive abilities were considered individually or as a set of simultaneous predictors. Thus Tenet 2 of CAD was supported by the predictiveness of the MCAT's *g*-factor, but challenged by the predictiveness of the MCAT's narrow cognitive abilities. Although this provides mixed support for CAD, it corresponds to Schmidt and Hunter's (1998) arguments that predictive validity is primarily driven by *g*, rather than narrow abilities. The findings concerning the stronger relations between the specific abilities in the low-*g* groups may potentially be explained by high-*g* individuals finding the tasks (i.e., course work associated with GPA) relatively easy, which may not tap the full range of abilities, thereby attenuating correlations.

6.3. Contributions and directions for future research

The results of this study ultimately supported Tenet 1 of CAD and provided mixed support for Tenet 2. Accordingly, depending on whether an individual had high- versus low-*g*, we observed that MCAT subtest *g*-factor loadings and residual variances differed substantially, as did the *g*-factor variances, and the predictiveness of the *g*-factor. Thus, the MCAT may demonstrate differential functioning across applicants' levels of *g*. Although medical schools may prioritize the acceptance of high-*g* individuals, recent research highlights the value of also considering non-cognitive attributes such as personality in selection (Goffin et al., 2011; McLarnon et al., 2017). If medical schools allow non-cognitive selection criteria to play key roles in their decisions, they may accept individuals from both the high- and low-*g* segments of the applicant pool. These individuals' respective MCAT scores may consequently function differently, in accordance with the differences uncovered by this study. Thus, the MCAT could be targeted for additional refinements to address the occurrence of CAD. This might improve selection decisions, as well as applicants' perceptions of justice regarding the selection process.

One advantage of this study was that data comprised four full cohorts of medical school applicants, representing the entire population of applicants within one sizable jurisdiction during those years. This supports the representativeness of our sample and the generalizability of the findings. Although dichotomization may have some limitations (i.e., range restriction, reduced power, negative correlations that may emerge [Jensen, 2003]), our use of the high- versus low-*g* design may be one of the few viable options to adequately assess research questions about differentiation (McLarnon & Carswell, 2013). Further, in comparing the high- and low-*g* thirds of the overall sample, these groups still contained over 2400 applicants, which afforded abundant statistical power. Future advances in moderated factor models may permit a more comprehensive analytical strategy to assess CAD. For the purposes of this study, the multi-group approach balanced the analytical needs of the research questions with the currently available statistical frameworks.

Additionally, most previous investigations of CAD have been restricted to low-stakes situations, whereas this study is distinctive because of the high-stakes scenario involved. Nonetheless, our results pertain to the MCAT, the cognitive performance criterion of GPA, and the use of medical school applicants. Additionally, testing tools such as the MCAT may include implicit range restriction, as test-takers representing the full-range of abilities measures may not take the test (very low-*g* individuals might not even consider applying to medical school). Thus, the low extreme group considered here may not objectively reflect low-*g* individuals in the population. However, this restriction in range afforded this study a conservative test of differentiation because individuals in the low- and high-*g* groups may not have been as strongly divergent in *g* relative to the general population. Cross-validation with other high-stakes tests (e.g., Graduate Records Exam, SAT), other performance criteria, and other samples are necessary directions for future research.

7. Conclusion

We examined cognitive ability differentiation by assessing the invariance of *g* as measured by the MCAT across low- and high-*g* individuals. This was accomplished using a large sample of medical school applicants. Although there was some degree of mixed evidence, it was generally found that the measurement properties of the MCAT were not invariant across low- and high-*g* test-takers. Additionally, we found that *g* and its narrow abilities demonstrated differential relations with GPA across low- and high-*g* groups, which partially supported differentiation. All things considered, our study contributes a unique examination of cognitive ability differentiation as reflected by the MCAT and data obtained from medical school applicants undergoing high-stakes testing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2017.11.005.

References

AAMC (1991). *Medical College Admission Test.* Washington: Association of American Medical Colleges.

Abad, F. J., Colom, R., Juan-Espinosa, M., & García, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31*, 157–166.

Bauer, D. J. (2016). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods, 22*, 507–526.

Blanch, A., García, H., Llaveria, A., & Aluja, A. (2017). The Spearman's law of diminishing returns in chess. *Personality and Individual Differences, 104*, 434–441.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research, 36*,

589–609.
Coyle, T., & Rindermann, H. (2013). Spearman's law of diminishing returns and national ability. *Personality and Individual Differences, 55*, 406–410.
Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*, 105–132.
Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for the low IQ groups. *Intelligence, 13*, 349–359.
Fogarty, G. J., & Stankov, L. (1995). Challenging the "law of diminishing returns". *Intelligence, 21*, 157–174.
Goffin, R. D. (2007). Assessing the adequacy of structural equation models: Golden rules and editorial policies. *Personality and Individual Differences, 42*, 831–839.
Goffin, R. D., Rothstein, M. G., Reider, M. J., Poole, A., Krajewski, H. T., Powell, D. M., ... Mestdagh, T. (2011). Choosing job-related personality traits: Developing valid personality-oriented job analysis. *Personality and Individual Differences, 51*, 646–651.
Hartmann, P., & Reuter, M. (2006). Spearman's "law of diminishing returns" tested with two methods. *Intelligence, 34*, 47–62.
Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31*, 95–105.
Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine, 80*, 910–917.
Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation.* Thousand Oaks, CA: Sage.
Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence, 23*, 45–57.
Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five Factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471–491.
McGaghie, W. C. (2002). Assessing readiness for medical education: Evolution of the Medical College Admission Test. *Journal of the American Medical Association, 288*, 1085–1090.
McLarnon, M. J. W., & Carswell, J. J. (2013). The personality differentiation by intelligence hypothesis: A measurement invariance investigation. *Personality and Individual Differences, 54*, 557–561.
McLarnon, M. J. W., Carswell, J. J., & Schneider, T. J. (2015). A case of mistaken identity? Latent profiles in vocational interests. *Journal of Career Assessment, 23*, 166–185.
McLarnon, M. J. W., Goffin, R. D., Schneider, T. J., & Johnston, N. G. (2016). To be or not to be: Exploring the nature of positively- and negatively-keyed personality items in high-stakes testing. *Journal of Personality Assessment, 98*, 480–490.
McLarnon, M. J. W., Rothstein, M. G., Goffin, R. D., Rieder, M. J., Poole, A., Krajewski, H. T., ... Mestdagh, T. (2017). How important is personality in the selection of medical school students? *Personality and Individual Differences, 104*, 442–447.
McNeish, D., An, J., & Hancock, G.R. The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment* (in press).
Murray, A. L., Dixon, H., & Johnson, W. (2013). Spearman's law of diminishing returns: A statistical artifact? *Intelligence, 41*, 439–451.
Muthén, L. K., & Muthén, B. (2012). *Mplus user's guide.* Los Angeles: Muthén & Muthén.
Nyborg, H., & Jensen, A. R. (2000). Black-white differences on various psychometrics tests: Spearman's hypothesis tested on American armed services veterans. *Personality and Individual Differences, 28*, 593–599.
Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence, 38*, 231–241.
Saklofske, D. H., Yang, J., Zhu, J. J., & Austin, E. J. (2008). Investigating Spearman's Law of Diminishing Returns in normative samples for the WISC-IV and WAIS-III. *Journal of Individual Differences, 29*, 57–69.
Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*, 347–363.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.
Spearman, C. E. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.