Intelligence 95 (2022) 101701

Contents lists available at ScienceDirect

Intelligence

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

Distributions of academic math-verbal tilt and overall academic skill of students specializing in different fields: A study of 1.6 million graduate record examination test takers

Jonathan Wai [a,*], Matthew H. Lee [b], Harrison J. Kell [c]

[a] University of Arkansas, 201 Graduate Education Building, Fayetteville, AR 72701, USA
[b] ACSI, P.O. Box 62249, Colorado Springs, CO 80962, USA
[c] ETS, 660 Rosedale Road, Princeton, New Jersey, 08541, USA

ARTICLE INFO

Keywords:
Academic achievement
Cognitive tilt
Field specialization
GRE
Math skill
Verbal skill

ABSTRACT

Using a sample of over 1.6 million scores of U.S. test takers on the Graduate Record Examination 2015–2020, this study broadly replicated prior findings going back over seven decades on overall academic skill and math-verbal tilt as a function of different field specialization. Individuals pursuing STEM degrees and STEM undergraduate backgrounds had stronger quantitative than verbal skills. Individuals pursuing arts/humanities degrees and arts/humanities undergraduate backgrounds had stronger verbal than quantitative skills. However, there were also differences regarding math-verbal tilt in the GRE relative to other samples. Academic skill patterns may be both a cause of or result of educational choices, and deeper consideration of these issues may ultimately have implications for expertise development for students who pursue fields such as the STEM or the arts/humanities.

1. Studying the distribution of academic skill across major fields: historical context

Do students who choose to major in different fields have different academic skills? If so, does the pattern of specific academic skills vary by field, and has this changed over time? Investigating these questions is worthwhile for many reasons, including understanding what fields different students choose to pursue, the diversity of skills across fields, and historical consistencies – and divergences – in the niches individuals with differing cognitive skillsets sort into. For example, academic skill patterns may be in part either a cause of educational choices into different areas (e.g., STEM or the humanities), or the result of educational choices. Thus, the study of academic skill or "tilt" patterns may have implications for the development of expertise (Hambrick, Campitelli, & Macnamara, 2017) in various domains. In this introduction we briefly review the literature on academic skill patterns using diverse measures starting in the 1950s, moving to the 1970s, and then the early 2000s prior to introducing the current analyses conducted in this study.

Roe (1951, 1953) conducted early studies of the cognitive skills of research scientists across different fields. Broader examinations across multiple academic major fields and occupations were detailed in work by Wolfle (1954) and Wolfle and Oxtoby (1952), who illustrated median scores on the Army General Classification Test as a function of college major in a sample of college graduates from 40 universities in 1946. These authors showed education, agriculture, and business and commerce tended to have lower average scores, whereas humanities, engineering, and physical sciences tended to have the highest scores, with social and biological sciences in the middle. Wolfle and Oxtoby (1952) also presented mean scores of 38,420 college seniors who took the Selective Service College Qualification Test in 1951, uncovering essentially the identical average skill pattern across majors.

2. Academic skill as a function of math-verbal "tilt" pattern by field

Fast forward to data gathered during the early 1970s, from which Wai, Lubinski, and Benbow (2009) used a stratified random sample of roughly 400,000 students who were tested in high school on verbal, math, and spatial skill and followed up well after they earned their terminal degrees. Wai et al. (2009, see Figure B1, p. 834) showed student overall academic skill by major, as well as the pattern of their verbal, math, and visuospatial skills within major as a function of earning a terminal bachelor's, master's, or PhD degree. The general pattern from the findings summarized in Wolfle and Oxtoby (1952) was

* Corresponding author.
E-mail addresses: jwai@uark.edu (J. Wai), hmatthewlee@gmail.com (M.H. Lee), hkell@ets.org (H.J. Kell).

https://doi.org/10.1016/j.intell.2022.101701
Received 23 August 2022; Received in revised form 23 September 2022; Accepted 25 September 2022

replicated. Beyond those findings, however, Wai et al. (2009) also showed that there appeared to be different "tilt" patterns of cognitive skills within majors, with intellectual tilt being the idea that, although individuals may have comparable overall general or composite cognitive skill, their specific skill strengths can vary and these patterns may forecast important choices that individuals make in terms of their educations, occupations, and creative endeavors (cf. Athey, Katz, Krueger, Levitt, & Poterba, 2007; Coyle, 2018a, 2018b; Kell, Lubinski, & Benbow, 2013; Lubinski, 2009; Makel, Kell, Lubinski, Putallaz, & Benbow, 2016; Park, Lubinski, & Benbow, 2007). These consistent results affirmed the practical value of specific reasoning skills, even in the face of criticism that they lack importance after taking into account general reasoning (e.g., Schmidt, 2012); whereas general reasoning may be ideal for predicting the magnitude of outcome individuals achieve (e.g., highest degree attained), specific reasoning skills – and their relative strength – are ideal for predicting the types of outcomes individuals achieve (e.g., field in which a degree is attained) (Kell et al., 2013; Lubinski, 2010).[1]

Although the centerpiece of Wai et al.'s (2009) analysis was a random sample drawn from the general population, it also featured an examination of verbal and math intellectual tilt patterns using data from Graduate Record Examinations (GRE) test-takers from 2002 to 2005. These findings replicated the major pattern observed in Wai and colleagues' general sample, along with in other general (e.g., Coyle, 2018b; Coyle, Snyder, & Richmond, 2015) and gifted (e.g., Kell et al., 2013; Makel et al., 2016) samples: Individuals with test scores from disciplines in science, technology, engineering and mathematics (STEM) (e.g., computer science, physical science) tended to have intellectual profiles typified by math scores exceeding verbal scores, whereas individuals pursuing degrees in the arts and humanities tended to have intellectual profiles of the opposite type.

3. Current study

This paper builds on the empirical literature from prior decades showing the pattern of academic skills by major field (e.g., Wai et al., 2009; Wolfle & Oxtoby, 1952). In recognition of the need for replication in the psychological sciences (e.g., Shrout & Rodgers, 2018; Wiggins & Christopherson, 2019), we focus on replicating and extending part of the findings from Wai et al. (2009), specifically those featuring GRE test-takers, using a more recent sample. We expand on Wai et al.'s (2009) analyses by both examining intellectual tilt patterns according to both GRE-takers' undergraduate majors and the highest degree they were pursuing (i.e., master's vs. doctoral), using GRE Quantitative (GRE-Q) and Verbal (GRE-V) scores to assess math and verbal skills, respectively, among U.S. test takers. In addition to seeking to replicate and extend Wai et al.'s (2009) findings, by focusing on individuals pursuing advanced degrees our results uniquely contribute to the literature on cognitive tilt patterns, which has typically drawn from the general population (e.g., Coyle, 2014, 2018a, 2018b; Coyle & Pillow, 2008; Coyle, Purcell, Snyder, & Richmond, 2014) or populations identified as intellectually gifted in early adolescence (e.g., Kell et al., 2013; Makel et al., 2016; Park et al., 2007).

4. Study methods

4.1. Participants

Data were provided by Educational Testing Service (ETS) for all administrations of the GRE General Test from early 2015 to early 2020. The sample consisted of 1,603,294 unique U.S. test-taking candidates who completed 1,981,222 GRE General Tests. See Table 1 for detailed descriptive breakdowns. Candidates were coded as U.S. citizens if they ever indicated "United States Citizen" for current citizenship status or "USA" for country of citizenship on the Background Information Questionnaire (BIQ) they completed during any GRE test administration. Unique test-takers by year ranged from a minimum of 191,587 in 2020 to a maximum of 324,619 in 2015. Undergraduate department and intended graduate major were reported on 1,059,005 and 1,981,163 test observations respectively.

4.2. Variables

Test scores. ETS provided scores on GRE-Q and GRE-V, which range from 130 to 170 in one-point increments. Analytical Writing (AW) is scored on a scale of 0–6, in half point increments. Altogether, there were 1,977,222 (99.8%) GRE-Q scores ($M = 150.18$, $SD = 7.91$) and 1,981,579 (99.8%) GRE-V scores ($M = 152.38$, $SD = 7.74$).

Tilt. Tilt was based on within-subject differences in GRE-Q and GRE-V scores on the GRE General Test. Tilt was calculated as the difference in reported GRE-Q and GRE-V scores for all observations for which neither GRE-Q nor GRE-V were missing. Positive scores (GRE-Q > GRE-V) indicated a quantitative tilt; negative scores (GRE-Q < GRE-V) indicated a verbal tilt. We were able to calculate tilt for 1,975,862 (99.7%) test observations ($M = -2.20$, $SD = 6.60$). For our analytic sample of U.S. test-takers, we standardized GRE-Q, GRE-V, and overall GRE (GRE-Q + GRE-V) scores separately so that each had a mean of 0 and a standard deviation of 1 across all years of analysis.

Undergraduate major. We coded test-takers' undergraduate majors using their BIQ responses and the four-digit codes corresponding to undergraduate major fields of study (Educational Testing Service, 2019). Department and major classifications were identified for 1,059,005 (53.5%) of all test observations.[2] Following Educational Testing Service (2019) classifications, we coded each test observation according to one of the following undergraduate majors: Life Sciences (351,220, 33.2% of the sample with identified undergraduate major), Physical Sciences (including Engineering; 183,224, 17.3%), Social and Behavioral Sciences (180,780, 17.1%), Humanities & Arts (80,823, 7.6%), Education (56,784, 5.4%), Business (47,535, 4.5%), and Other or Undecided (158,639, 15.0%). Given that the GRE General Test has only recently begun to be accepted by law schools, we excluded data from individuals who indicated their undergraduate major was Law.

Intended graduate major and degree. Intended graduate major was available for 1,981,163 test observations (99.997%). We coded intended major using Educational Testing Service (2019) department and major classifications: Life Sciences (663,325, 33.5%), pHysical Sciences (including Engineering; 265,671, 13.4%), Social and Behavioral Sciences (278,047, 14.0%), Humanities & Arts (87,706, 4.4%), Education (139,578, 7.1%), Business (122,558, 6.2%), and Other or Undecided (424,278, 21.4%). Given the GRE General Test has only recently begun to be accepted by law schools, we excluded data from individuals who indicated their intended graduate major was Law.

Test-takers were also asked to indicate their eventual graduate education objectives. We focused on a sample of 1,608,344 test observations in which the graduate objective was reported as "Master's (M.A., M.S., M.Ed.)," "Intermediate (e.g., Graduate Certificate, Specialist),"

[1] The fact that specific reasoning strengths have repeatedly been found to forecast areas in which people achieve important outcomes (e.g., highest degree attainment, production of creative works) suggests specific reasoning skills are substantive variables, rather than merely being the products of measurement error in estimating general reasoning. We thank an anonymous reviewer for drawing our attention to this interpretation.

[2] Table 4 displays descriptive statistics for test-takers who did and did not report their undergraduate majors. We did not observe major differences between reporters and non-reporters. Reporters were slightly higher achievers, had slightly less verbal tilt, were younger, and more likely to pursue graduate careers in the Life Sciences, Physical Sciences, Humanities, less likely to pursue Education, Business, or be Undecided. Reporters were also more likely to pursue a doctoral than master's degree.

Table 1
Descriptive Statistics by Year.

	2015	2016	2017	2018	2019	2020
	(1)	(2)	(3)	(4)	(5)	(6)
N GRE General Tests	366,736	367,803	365,275	345,562	320,117	215,729
N Unique test-takers	324,619	323,754	320,240	302,498	280,165	191,587
N Unique test-takers overall	1,603,294					
Intended graduate major						
Life Sciences	120,158	120,715	119,745	116,316	109,108	77,283
Physical Sciences	47,347	49,005	49,354	47,326	44,577	28,062
Social and Behavioral Sciences	53,839	52,182	50,419	46,672	44,405	30,530
Humanities & Arts	19,283	17,617	16,758	14,817	12,654	6577
Education	25,198	27,062	28,046	25,150	21,900	12,222
Business	20,446	20,568	20,873	21,912	21,681	17,078
Other or Undecided	80,447	80,640	80,065	73,361	65,790	43,975
Degree Objective						
Master's	161,816	176,574	184,789	171,895	153,374	105,709
Doctorate	106,374	115,363	121,754	116,202	106,606	68,892
Undergraduate Department						
Life Sciences	67,333	65,739	64,432	60,130	55,035	38,551
Physical Sciences	34,004	35,094	35,014	31,837	29,339	17,936
Social and Behavioral Sciences	37,103	35,336	33,976	29,695	27,443	17,227
Humanities & Arts	18,345	17,004	15,528	12,927	11,041	5978
Education	11,097	11,322	11,102	9853	8508	4902
Business	8952	8731	8721	8422	7659	5050
Other or Undecided	29,107	29,147	30,926	28,093	25,097	16,269
Scores						
GRE-Q	150.13	150.11	150.10	150.12	150.19	150.63
SD	7.76	7.81	7.87	7.96	8.01	8.12
GRE-V	152.49	152.41	152.39	152.31	152.26	152.41
SD	7.63	7.65	7.76	7.79	7.81	7.83
GRE-AW	3.83	3.83	3.86	3.87	3.88	3.96
SD	0.77	0.78	0.80	0.80	0.80	0.81

"Doctorate (e.g., Ph.D., Ed.D.)," "M.B.A.," or "Specialized master's in business." Test observations indicating "Master's," "Intermediate," "M. B.A.," or "Specialized master's in business" were coded as pursuing a Master's (n = 954,157 test observations), while test observations indicating "Doctorate" were coded as pursuing a Doctorate (n = 635,191 test observations).

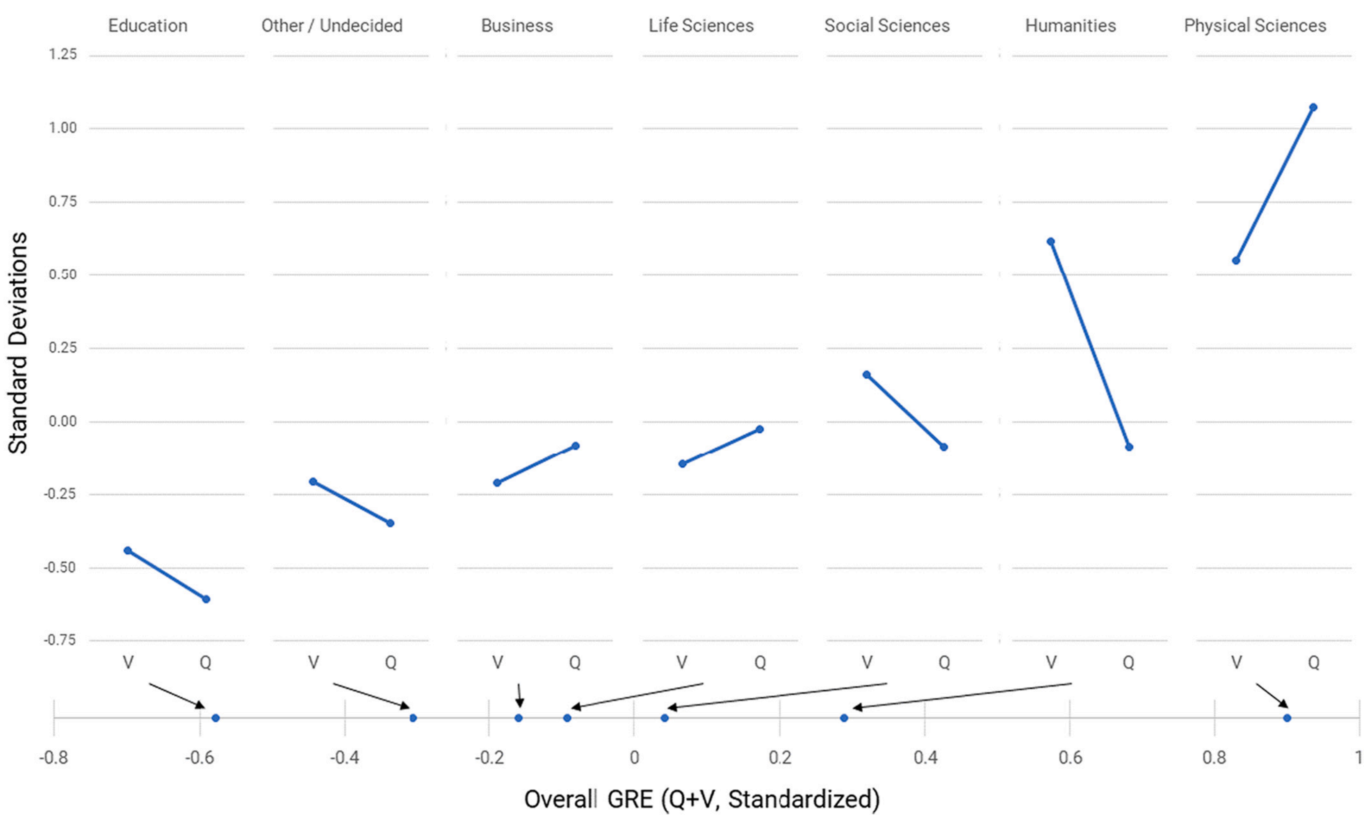

Fig. 1. Math-verbal tilt by undergraduate field of study.

4.3. Analyses

For undergraduate major, we estimated the average GRE-Q, GRE-V, and overall GRE (GRE-Q + GRE-V) scores within each field of study, pooled across all years in the study sample. For intended graduate major, we estimated the average GRE-Q, GRE-V, and overall GRE scores within each department by graduate objective, pooled across all years in the sample.

5. Results

In Fig. 1, we present GRE-V and GRE-Q scores, which were standardized across the entire sample, by test-takers' undergraduate department, plotting the average standardized overall GRE score below. See Table 2 for detailed descriptive statistics. Test-takers representing three undergraduate department classifications (Education, Social Sciences, and Humanities) together with test-takers undecided about their undergraduate major at the time of testing had a verbal tilt (GRE-V > GRE-Q), whereas test-takers representing the remaining three undergraduate department classifications (Business, Life Sciences, and Physical Sciences) had a quantitative tilt (GRE-Q > GRE-V). Education majors had the lowest overall GRE scores, nearly three-fifths of a standard deviation below the mean on average, whereas physical science majors had the highest overall GRE scores, roughly nine-tenths of a standard deviation above the mean on average. Physical science majors had the steepest quantitative tilt, with over half a standard deviation more favorable performance on GRE-Q relative to GRE-V. Humanities majors, the second highest overall performing group, had the steepest verbal tilt, with nearly three-quarters of a standard deviation more favorable performance on the GRE-V relative to GRE-Q.

In Fig. 2, we present standardized GRE-V and GRE-Q scores by test-takers' intended graduate major, plotting the average standardized overall GRE score below. See Table 3 for more detailed descriptive information. Overall GRE scores were calculated by pooling the scores of test-takers intending to pursue a master's or doctoral degree. Test-takers intending to pursue a master's are plotted in solid lines, and test-takers intending to pursue a doctorate are plotted in dashed lines.

Performance patterns by intended graduate major largely mirrored those observed across undergraduate fields of study. Test-takers intending to pursue graduate studies in Education, Social Sciences, and Humanities & Arts, along with test-takers undecided about their intended graduate major at the time of testing had a verbal tilt (GRE-V > GRE-Q). Those who intended to pursue graduate studies in Life Sciences and Physical Sciences had a quantitative tilt (GRE-Q > GRE-V). Aspiring Physical Science graduate students again had the highest overall performance, roughly four-fifths of a standard deviation above the mean on average, and aspiring Education graduate students again had the lowest overall performance, roughly two-fifths of a standard deviation below the mean on average. Aspiring Physical Science graduate students again had the steepest quantitative tilt, roughly half a standard deviation favorable performance on GRE-Q relative to GRE-V. Aspiring Humanities & Arts graduate students again had the steepest verbal tilt, roughly three-quarters of a standard deviation favorable performance on GRE-V relative to GRE-Q. Test-takers intending to pursue a doctorate had higher GRE-Q and GRE-V scores than test-takers intending to pursue a master's, with an exception. Aspiring Business graduate students' performance on GRE-Q was identical on average, regardless of their intent to pursue a master's or a doctorate.

Findings for intended field of graduate study versus undergraduate major differed in two ways. First, the average GRE scores of test-takers aspiring to graduate careers in Business were higher than those pursuing advanced degrees in Life Sciences, whereas this relative position was reversed when GRE scores were ordered according to undergraduate field of study. Second, test-takers with undergraduate majors in Business manifested quantitatively-tilted profiles, whereas test-takers pursuing doctoral degrees in Business manifested essentially flat cognitive profiles.

In a set of additional analyses we attempted to more rigorously quantify and study flat intellectual profiles. We did this by deriving the standard error of measurement (SEM) of the difference between GRE-Q and GRE-V scores (3.34 points) and then defining individuals whose tilt score did not exceed this value as having flat profiles. Table 5 depicts the

Table 2
GRE Scores by Intended Graduate Field.

	Education	Other or Undecided	Life Sciences	Business	Social Sciences	Humanities & Arts	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Doctorate							
V	152.00	152.79	151.41	152.66	154.27	158.19	157.11
	7.52	7.88	6.93	8.04	7.27	6.83	6.91
Q	146.69	148.53	150.34	150.45	149.79	149.26	158.94
	7.32	7.75	6.50	8.36	7.38	7.46	6.80
AW	3.91	3.90	3.84	3.88	4.08	4.31	4.12
	0.79	0.82	0.71	0.85	0.77	0.77	0.75
Q-V	−5.31	−4.26	−1.07	−2.21	−4.48	−8.93	1.84
	6.40	6.79	5.53	6.51	5.90	6.32	6.01
V + Q	298.69	301.32	301.76	303.12	304.06	307.46	316.06
	13.39	14.08	12.24	15.04	13.41	12.83	12.27
N	38,737	79,760	246,103	7324	127,486	41,194	93,323
Master's							
V	149.54	150.52	150.39	152.16	152.23	155.14	154.95
	7.84	7.89	6.74	7.72	7.88	7.46	7.17
Q	146.09	147.35	148.59	150.45	148.11	147.58	157.04
	7.20	7.44	6.33	7.43	7.43	7.17	6.93
AW	3.65	3.71	3.78	3.86	3.87	4.02	3.93
	0.84	0.81	0.72	0.83	0.82	0.80	0.76
Q-V	−3.45	−3.18	−1.81	−1.72	−4.12	−7.56	2.10
	6.16	6.24	5.53	6.10	5.89	6.37	6.37
V + Q	295.64	297.88	298.99	302.62	300.34	302.72	312.00
	13.74	13.99	11.84	13.86	14.10	13.17	12.58
N	74,976	251,735	288,917	88,292	102,828	30,900	113,686
Overall							
V + Q	296.68	298.61	300.26	300.85	302.34	305.42	313.85
	13.69	14.06	12.10	14.03	13.83	13.19	12.60

Note. Standard deviations reported below means.

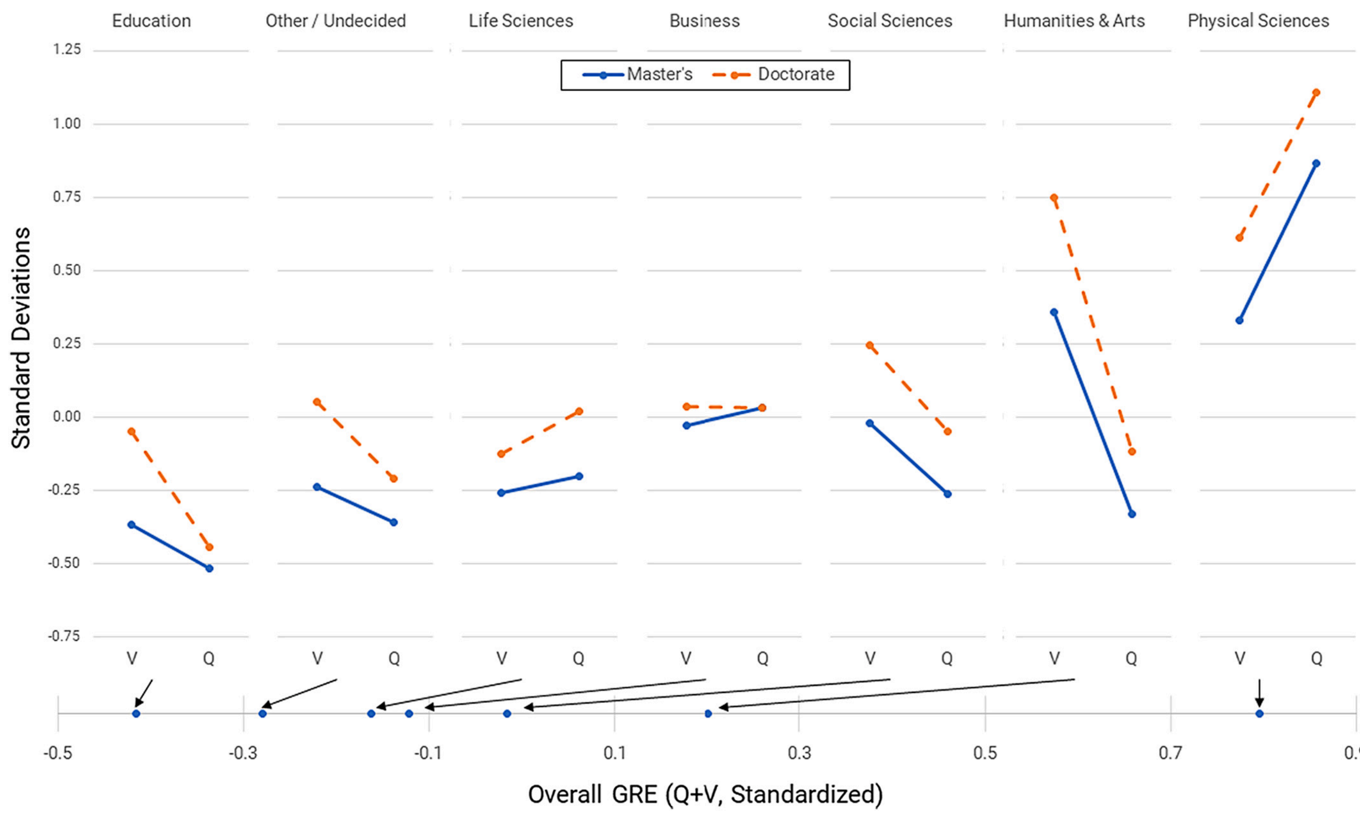

Fig. 2. Math-verbal tilt by intended graduate major.

Table 3
GRE Scores by Undergraduate Department.

	Education	Other or Undecided	Business	Life Sciences	Social Sciences	Humanities & Arts	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
V	148.98	150.79	150.76	151.26	153.64	157.14	156.64
	7.33	7.82	7.37	6.76	7.62	7.31	6.92
Q	145.40	147.45	149.54	149.99	149.51	149.51	158.67
	6.81	7.39	7.23	6.49	7.48	7.37	6.80
AW	3.64	3.76	3.72	3.83	4.00	4.20	4.06
	0.81	0.81	0.81	0.71	0.79	0.78	0.75
Q-V	−3.57	−3.35	−1.22	−1.28	−4.13	−7.63	2.02
	6.22	6.20	5.92	5.43	5.79	6.24	5.95
V + Q	294.38	298.24	300.30	301.25	303.16	306.66	315.31
	12.70	13.89	13.34	12.10	13.94	13.29	12.37
N	56,651	157,860	47,452	350,465	180,520	80,648	183,096

Note. Standard deviations reported below means.

results of these analyses. We found that test-takers intending to earn graduate degrees in the Physical Sciences, Social Sciences, and Business did not manifest tilt exceeding the SEM of the difference between the two test scores, wheras individuals intending to pursue graduate pathways in Education were the least likely to display flat profiles. In total, 40% of the sample had no tilt, wheras 18% had a quantitative tilt and 42% had a verbal tilt.

6. Discussion

We replicated the key pattern observed in prior research on tilt in diverse types of samples: Individuals pursuing STEM degrees (and with STEM undergraduate backgrounds) tended to have stronger quantitative than verbal skills and individuals pursuing arts/humanities degrees (and with undergraduate backgrounds in the arts/humanities) tended to have stronger verbal than quantitative skills. This is the same pattern that was found among Wai et al.'s (2009) prior study of GRE test-takers, along with studies featuring samples drawn from the general population (e.g., Coyle et al., 2014, 2015; Coyle, 2018a, 2018b; Wai et al., 2009) and intellectually talented samples (e.g., Park et al., 2007). Trends in total GRE scores across disciplines also broadly replicated those observed in Wai et al.'s (2009) general sample, with the disciplines having the highest overall scores being in STEM and the lowest overall scores being in education. Additionally, test takers intending to earn a doctorate had higher overall test scores than test takers intending to earn a master's degree, also replicating Wai and colleagues' findings.

More nuanced findings from Wai et al.'s (2009) study of GRE test-takers also replicated, such as individuals pursuing degrees in the social sciences tending to have verbally tilted profiles. Nonetheless, there were a few divergences. For example, test-takers in the current sample pursuing advanced degrees in education had relatively stronger verbal skills, whereas individuals pursuing advanced degrees in education in Wai et al.'s (2009) showed the opposite profile.

There were also some unique and notable differences in the GRE

Table 4
Descriptive Statistics by Undergraduate Department Reporting.

	Reported UGD		Did not report UGD	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
n	1,059,005		922,217	
Demographics				
Age	24.64	5.94	25.75	6.65
Testing accommodation	0.01	0.09	0.01	0.11
Race/Ethnicity				
American Indian	0.01	0.07	0.00	0.07
Asian	0.07	0.25	0.08	0.27
Black	0.07	0.26	0.08	0.28
Mexican	0.03	0.18	0.03	0.18
Pacific Islander	0.00	0.05	0.00	0.05
Puerto Rican	0.01	0.10	0.01	0.10
Hispanic	0.05	0.21	0.05	0.22
White	0.60	0.49	0.51	0.50
Other	0.04	0.19	0.04	0.20
GRE Scores				
Q	150.73	7.96	149.56	7.80
V	152.83	7.64	151.86	7.81
AW	3.90	0.78	3.83	0.81
Q-V	−2.11	6.37	−2.30	6.85
V + Q	303.57	14.24	301.42	14.01
Intended Graduate Studies				
Life Sciences	0.34	0.47	0.33	0.47
Physical Sciences	0.15	0.36	0.11	0.31
Social Sciences	0.14	0.35	0.14	0.35
Humanities & Arts	0.05	0.21	0.04	0.20
Education	0.07	0.26	0.07	0.26
Business	0.05	0.23	0.07	0.26
Other / Undecided	0.20	0.40	0.23	0.42
Doctorate	0.40	0.49	0.39	0.49
Master's	0.60	0.49	0.61	0.49

Note. UGD = Undergraduate department.

sample compared to the Project Talent sample, which was drawn from the general population. For example, in the GRE sample, both undergraduate humanities majors and those intending to pursue humanities graduate degrees earned higher test scores, overall, than those pursuing the life sciences, with the opposite pattern observed in Project Talent's data. Additionally, among GRE test-takers, the profiles of those pursuing the social sciences were verbally-tilted, whereas the profiles of Project Talent participants in the social sciences were quantitatively-tilted. Interestingly, GRE test-takers pursuing the humanities evinced strongly verbally-tilted profiles – yet in the Project Talent sample individuals with master's degrees in the humanities tended to have slightly higher math than verbal skills, with this quantitative tilt being substantially larger among holders of doctoral degrees in the humanities. The fact that the GRE sample were those "aspiring" to a particular graduate major/degree may have contributed to some discrepancies between findings across these two samples.

Our finding that a relatively large number of test-takers manifested flat cognitive profiles broadly replicates research in the gifted field. For example, investigators of a study conducted over 25 years ago (Achter, Lubinski, & Benbow, 1996) using a sample of young adolescents scoring in the top 1% of cognitive reasoning found that 54% of their sample manifested flat intellectual profiles.

7. Discrepancies in findings from prior research and possible limitations

We cannot pinpoint causes of these discrepancies using the current dataset, but can hypothesize about possible reasons for the differences and consider finding ways to investigate these possibilities in future research. One explanation for differences in these findings relative to prior results is that the Project Talent sample was representative of the general U.S. population, whereas the population of GRE test-takers represents a relatively self-selected sample of individuals who have earned, or are in the process of earning, four-year degrees. The GRE test-taker sample also differs from prior samples in several other ways, each of which individually – or combined – may have contributed to the discrepant findings. First, prior data used to investigate cognitive tilt were gathered in the 1960s (Project Talent) and 1970s–1980s (Study of Mathematically Precocious Youth), whereas GRE test-takers' scores were gathered in a much more contemporary time period (2015 to 2020). Second, cognitive tests were administered much earlier in individuals' developmental trajectories (early adolescence and high school) than in the current sample. Third, prior research on tilt generally relied on obtained advanced degrees, but the GRE test-taker sample features only intended advanced degrees. Another issue is self-selection into programs that do or do not accept GRE scores. Finally, prior work relied on test scores produced under relatively low-stakes conditions, whereas the GRE is typically only taken in high-stakes circumstances. For example, examinee motivation to perform well on the GRE overall or on a particular subtest of the GRE for admission to a particular program may have played a role. It is also possible that more detailed investigation of subgroup differences (e.g., domestic vs. international test-takers) may shed light on variation in findings. Historically, academic skill by major may have somewhat shifted over the years, perhaps as a reflection of changes in U.S. society or culture.

8. Implications and conclusions

For whatever reason, the average math and verbal skill levels of students by academic field has remained remarkably robust across the last seven or more decades in the U.S. Given this general pattern is found using different types of standardized cognitive and achievement tests on a wide variety of samples at very different points in historical time suggests the patterns are meaningful. Cognitive aptitudes research suggests that overall cognitive skill and math-verbal skill both contribute meaningfully to educational, occupational, and broader life trajectories. Thus, talent selection and development discussions surrounding different educational and occupational niches should likely be informed by this work. For example, what are the implications that the education field remains consistently at the bottom (Eide, Goldhaber, & Brewer, 2004)? What are the implications that hard STEM fields tend to be consistently at the top (Wai et al., 2009)? Should there be efforts to change the academic skill distributions by major field in the U.S.? If so, how might one try to go about that given society appears to have greatly changed over the past several decades and yet the academic skill

Table 5
Percentage of Test-Takers within One standard Error of Measurement of Tilt.

	Education	Other / Undecided	Life Sciences	Business	Social Sciences	Humanities & Arts	Physical Sciences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Q Tilt	11.3%	14.0%	18.3%	9.3%	3.6%	1.2%	23.2%
% No Tilt	35.4%	37.6%	46.9%	69.7%	61.6%	37.4%	69.0%
% V Tilt	53.3%	48.4%	34.8%	20.9%	34.8%	61.4%	7.8%
N	139,228	422,372	661,587	122,314	277,643	87,358	265,305

Note. "Q tilt" was defined as test-takers whose GRE-Q score exceeded GRE-V score by at least one standard error of measurement (3.34 points). "V tilt" was defined as test-takers whose GRE-V score exceeded GRE-Q score by at least one SEM. "No tilt" was defined as test-takers whose GRE-Q and GRE-V scores were within one SEM. GRE-Q and GRE-V scores correlated 0.6437 for test observations with valid GRE-Q and GRE-V scores.

distributions by academic domain have not largely changed? At the same time the general pattern has remained robust, the findings within the GRE sample in our study here were not exactly the same for various major fields, showing that things can and do change. Why have some things changed but not others? Perhaps future research could investigate these issues helping to understand historical shifts in selection in various areas and also add to the literature of our understanding of cognitive aptitudes research.

Cognitive skill sorting into various academic domains also provides a window into understanding from one measurement lens what domains U.S. culture may value at any given time. We believe this study on over 1.6 million GRE test takers among a contemporary sample provides an important lens on understanding how highly educated and talented students are choosing what domains to pursue—and develop their expertise in (e.g., Hambrick et al., 2017)—and should be carefully tracked in the future.

Declaration of Competing Interest

None.

Data availability

The data that has been used is confidential.

Acknowledgement

This work was conducted in part due to a grant from the ETS titled "Historical Trend Analysis of GRE Scores by Test Taker Characteristics, Test Taking Frequency, and Tilt or Score Patterns."

References

Achter, J. A., Lubinski, D., & Benbow, C. P. (1996). Multipotentiality among the intellectually gifted: "it was never there and already it's vanishing.". *Journal of Counseling Psychology, 43*(1), 65–76. https://doi.org/10.1037/0022-0167.43.1.65

Athey, S., Katz, L. F., Krueger, A. B., Levitt, S., & Poterba, J. (2007). What does performance in graduate school predict? Graduate economics education and student outcomes. *American Economic Review, 97*, 512–520. https://doi.org/10.1257/aer.97.2.512

Coyle, T. R. (2014). Predictive validity of non-*g* residuals of tests: More than *g*. *Journal of Intelligence, 2*, 21–25. https://doi.org/10.3390/jintelligence2010021

Coyle, T. R. (2018a). Non-*g* factors predict educational and occupational criteria: More than *g*. *Journal of Intelligence, 6*(3), 43. https://doi.org/10.3390/jintelligence6030043

Coyle, T. R. (2018b). Non-*g* residuals of group factors predict ability tilt, college majors, and jobs: A non-*g* nexus. *Intelligence, 67*, 19–25. https://doi.org/10.1016/j.intell.2017.12.003

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing *g*. *Intelligence, 36*, 719–729. https://doi.org/10.1016/j.intell.2008.05.001

Coyle, T. R., Purcell, J. M., Snyder, A. C., & Richmond, M. C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence, 46*, 18–24. https://doi.org/10.1016/j.intell.2014.04.008

Coyle, T. R., Snyder, A. C., & Richmond, M. C. (2015). Sex differences in ability tilt: Support for investment theory. *Intelligence, 50*, 209–220. https://doi.org/10.1016/j.intell.2015.04.012

Educational Testing Service. (2019). Department & major field codes. https://www.ets.org/s/gre/pdf/dept_major_field_codes.pdf.

Eide, E., Goldhaber, D., & Brewer, D. (2004). The teacher labour market and teacher quality. *Oxford Review of Economic Policy, 20*(2), 230–244. https://doi.org/10.1093/oxrep/grh013

Hambrick, D. Z., Campitelli, G., & Macnamara, B. N. (2017). *The science of expertise*. UK: Routledge.

Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science, 24*(5), 648–659. https://doi.org/10.1177/0956797612457784

Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavior Genetics, 39*(4), 350–358. https://doi.org/10.1007/s10519-009-9273-0

Lubinski, D. (2010). Neglected aspects and truncated appraisals in vocational counseling: Interpreting the interest–efficacy association from a broader perspective: Comment on Armstrong and Vogel (2009). *Journal of Counseling Psychology, 57*(2), 226–238. https://doi.org/10.1037/a0019163

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science, 27*(7), 1004–1018. https://doi.org/10.1177/0956797616644735

Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science, 18*(11), 948–952. https://doi.org/10.1111/j.1467-9280.2007.02007.x

Roe, A. (1951). A study of imagery in research scientists. *Journal of Personality, 19*(4), 459–470. https://doi.org/10.1111/j.1467-6494.1951.tb01507.x

Roe, A. (1953). *The making of a scientist*. New York: Dodd, Mead & Company.

Schmidt, F. L. (2012). Cognitive tests used in selection can have content validity as well as criterion validity: A broader research review and implications for practice. *International Journal of Selection and Assessment, 20*(1), 1–13.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817–835. https://doi.org/10.1037/a0016127

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology, 39*(4), 202–217. https://doi.org/10.1037/teo0000137

Wolfle, D. (1954). *America's resources of specialized talent*. Report of the Commission on Human Resources and Advanced Training. Harper and Brothers.

Wolfle, D., & Oxtoby, T. (1952). Distributions of ability of students specializing in different fields. *Science, 116*(3013), 311–314. https://doi.org/10.1126/science.116.3013.311