A Pilot Study of Race, Ethnic, Gender, and Parental Education Level Differences on the Naglieri General Ability Tests: Verbal, Nonverbal, and Quantitative

Selvamenan, M., Paolozza, A., Solomon, J., Naglieri, J. A., & Schmidt, M. T. (2022)

Abstract: The present study was conducted to examine the relationships between general intelligence test scores and race, ethnicity, gender, and parental education using the Naglieri General Ability Tests (Naglieri et al., 2021) for three samples that closely match the U.S. population. Few differences were found on the preliminary versions of the Verbal (Naglieri & Brulles, 2021) (N = 2,482), Quantitative (Naglieri & Lansdowne, 2021) (N = 2,369), and Nonverbal (Naglieri, 2021) (N= 2,383) Naglieri General Ability Tests. These initial findings suggest that this approach to measuring general ability may ultimately have utility for equitable identification of students from diverse backgrounds and possible inclusion in gifted educational programs.

Introduction

Achieving equitable representation of students from diverse cultural, racial, ethnic, linguistic, or socioeconomic backgrounds, or from those who have had limited or different educational experiences, in gifted programming is an ongoing multidimensional issue (Brulles, et al., 2022; Gentry et al., 2019; Hodges et al., 2018; List & Dykeman, 2019; Peters & Engerrand, 2016; Wright et al., 2022). For example, Black and Hispanic students often have limited educational opportunity which can be related to limited outcomes (Castellano & Frasier, 2010; Ford, 2013; Gentry et al., 2019; Office for Civil Rights, 2012). According to the Office for Civil Rights (OCR; 2012) report, Black students were underrepresented by 50% in gifted education and Hispanic students by 40%. A recent meta-analysis revealed that the probability of gifted identification for Black, Hispanic, and Native American students was 66% less than for Asian and White students (Hodges et al., 2018). These findings are widespread throughout the nation, as shown by Gentry et al., (2019) and Yoon and Gentry (2009), who found that Hispanic, Black, and Native American students were underrepresented in gifted programs in nearly all U.S. states. The 2018–2019 State of the States in Gifted Education report (Rinn et al., 2020) provided data from 23 states and found that Black and Hispanic students encompassed 7.1 and 10.4%, respectively, of the total gifted population, and American Indian or Alaska Native students encompassed 1.6%. Given the need for equitable rates of identification, the goal of this study was to explore the possible utility of newly configured verbal, nonverbal, and quantitative tests across several demographic categories.

In addition to race and ethnicity, socioeconomic status (SES), commonly measured by indicators such as family income, parental education, and/or parental occupation, is another factor commonly associated with disproportionalities in gifted identification (List & Dykeman, 2019). Rinn, Mun, and Hodges (2020) also report rates of gifted identification for students who receive free/reduced lunch that were as low as 3% in some states even though students eligible for free/reduced lunch make up 52% of the student population (De Brey et al. 2021). Similarly, there is underrepresentation of English Language Learners (ELL). Rinn et al. (2020) reported that ELL students encompassed only 1.7% of the gifted population in 16 states, even though they comprise about 10% of the US student population.

The 2018–2019 State of the States in Gifted Education (Rinn, et. al., 2020) reported that males and females were closely represented in gifted education programs. This is consistent with small differences between the sexes reported for ability tests often used in the identification of gifted students. For example, Lohman reported that the largest differences on the Cognitive Abilities Test (CogAT; Lohman, 2012) for the standardization sample (*N* = 65,630) occurred at the "Primary level where females excel on both the Verbal and Nonverbal batteries by 2.6 and 2.0 SAS points, respectively. Thereafter the differences are all quite small… ranging from 0.1 to1.5 SAS points (p. 101)". Similarly, Rojahn and Naglieri (2006) studied sex differences for 79,780 children and adolescents from Kindergarten to Grade 12 who participated in

the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997) standardization. They found NAI score gender differences ranged from a 1.1-point advantage for males at NNAT level A to a 1.3- point advantage for females at NNAT level B. No significant NAI differences were found for NNAT levels C and D. Rojahn and Naglieri (2006) concluded that the small differences have little to no practical importance. Although there are some differences in the magnitude of the differences between the sexes reported by Lohman and Naglieri, the results for the CogAT and the NNAT are similar.

The disproportionalities in gifted enrollment rates are thought to be related, in part, to identification methods that fail to identify students from diverse populations (Brulles, et al., 2022; List & Dykeman, 2019). Several authors attribute the problem to measures that focus on exhibited talent (i.e., academic achievement; Brulles & Naglieri, 2022; Borland, 2009; Ford, 1998; Frasier et al, 1995; Pfeiffer, 2012; Wasserman & Becker, 2000) rather than high intellectual ability. Common identification methods across the nation often include intelligence test scores with a set cut-off to be identified as gifted (Robinson, 2005; Worrell, 2009). The most used individually and group administered intelligence tests have been shown to yield lower scores for students of color (Brulles et al., 2022; Naglieri & Otero, 2018; Pfeiffer, 2012; Wasserman & Becker, 2000). The verbal and quantitative content on the most widely used intelligence tests have been criticized for posing an obstacle for underrepresented groups who may not have had the opportunity to acquire the knowledge required to answer the questions included in these tests and therefore they fail to earn high enough scores to be accepted into gifted programs (Hodges et al., 2018; Brulles & Naglieri, 2022). This issue is closely related to the topic of fair and equitable assessment.

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) state that test content (prior knowledge), test context (complexity and comprehension demands of test instructions) and opportunity to learn (inequities in school resources) pose threats to fair and valid interpretation of test scores. The standards further explain that "opportunity to learn is a fairness issue when an authority provides differential access to opportunity to learn for some individuals and then holds those individuals who have not been provided that opportunity accountable for their test performance" (p. 57). In such instances, students' low scores may accurately reflect what they know and can do, so that, technically, the interpretation of the test results …may not be biased. However, it may be considered unfair to severely penalize students for circumstances that are not under their control" (p. 57). Thus, for students who have had limited opportunities to learn the content in, for example an intelligence test used as part of the process for access to gifted education, that test may be considered unfair because it penalizes students for not having the knowledge needed to answer the questions even if the norming data do not demonstrate psychometric test bias.

Measurement of ability using verbal test content also presents a dilemma because there is a bidirectional relationship between reading and improved scores on the verbal sections of intelligence tests (Ramsden et al., 2013). Skilled readers experience significant growth in Verbal scores, and the opposite is true for poor readers. For a student who has had opportunities to learn, verbal tests do not present much of an obstacle. In contrast, tests that demand knowledge present potential problems for those who have had limited educational opportunity, especially those from diverse populations.

To illustrate, test questions that demand academic knowledge of words are often found on intelligence tests as well as achievement tests (e.g., Woodcock-Johnson IV Oral Vocabulary subtest on the Tests of Cognitive Abilities (Schrank, McGrew & Mather, 2014) and Reading Vocabulary-Synonyms subtest on the Woodcock-Johnson-IV Tests of Achievement (Schrank, McGrew & Mather, 2014). When a vocabulary test is used to measure intellectual ability, the obtained score reflects what a student has learned *and* their ability to arrive at the answer by reasoning (Harris et al., 2007). A similar issue exists when verbal tests are used to measure both verbal reasoning and general knowledge taught in school and acquired from an enriched environment (e.g., Cognitive Abilities Test 7th Ed.; CogAT7; Lohman, 2011). The similarity in content (a) creates criterion contamination when studying the validity of an ability test; (b) may confound the measurement of ability with opportunity to learn; and (c) introduces a possible threat to the validity of scores for some examinees.

The problem introduced by tests that demand knowledge has been known for at least 100 years as described by Yoakum and Yerkes (1920) in their book *Army Mental Testing*. They noted that "Men who fail alpha [verbal and quantitative tests] are sent to beta [nonverbal tests] in order that injustice by reason of relative unfamiliarity with English may be avoided" (p. 19). Similarly, Pintner (1923) wrote "A good intelligence test must avoid as much as possible anything that is commonly learned by the subjects tested"

(p. 61). Today, the extent to which an intelligence test demands acquired knowledge likely plays a role in identification of gifted students from diverse populations and therefore poses a potential threat to equitable assessment.

The role of intelligence tests that demand knowledge and achievement tests was examined in recent litigation in *McFadden v. Board of Education for Illinois School District U-46* (2013). Even though about 40% of the students in School District U-46 were Hispanic, only 2% of students in the district's mainstream elementary school gifted programs were Hispanic. Admission to the "mainstream program run by the district known as SWAS (school within a school)" (p. 22) involved high achievement in reading and math test (MAP) "which plaintiffs' witnesses credibly demonstrated favored children with higher verbal skills and disfavored Minorities. The court ruled that the use of tests that demanded knowledge of English contributed to the underrepresentation of Hispanic students in the mainstream gifted program, writing that "gifted children for whom English is a second language would likely score lower on a [verbal] test than the nonverbal, culturally neutral test." (Naglieri, 2011, p. 24). It is important to note, however, that including a nonverbal test as part of the assessment process did not automatically solve the problem of underrepresentation (Lee et. al, 2021). Similarly, Brulles et. al. (2022) described identification of gifted students as a multi-dimensional task that required analysis of the way the entire assessment procedure was designed. Educational professionals who are responsible for identifying gifted students need to select and apply tests carefully to ensure that diverse populations who have been traditionally underrepresented have an equal opportunity to participate (Ford, 2013).

Some have criticized nonverbal tests of ability as being too narrow in focus and for not assessing verbal and quantitative reasoning given their relevance to regular and gifted education programs (Lohman, 2005). While it is often recommended to assess multiple content areas (e.g., verbal, nonverbal, and quantitative) for gifted identification (McClain & Pfeiffer, 2012), the amount of knowledge and language required in some verbal and quantitative tests, as well as lengthy test directions, and response format, could be problematic for students who have had limited educational experiences or other early opportunities to learn. As noted above, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) state that the complexity and comprehension demands of test instructions poses a threat to equitable assessment. Evidence for this assertion has been found, for example, by Gill et al. (2012) who reported that difficulty following directions can be related to underlying problems with grammar and semantic aspects of language. Engle, Carullo, and Collins (1991) found that students' ability to recall directions presented orally was related to their working memory capacity. Additionally, Abedi and Lord (2001) reported that language comprehension skills affect student performance on math tests. These findings suggest that the instructions for tests of general intellectual ability that use verbal, quantitative, and nonverbal content should be constructed in a way to address these concerns.

The Present Study

The overarching goal of this study was to determine if general ability tests with verbal, nonverbal and quantitative (a) questions that were explicitly written in a manner that reduces the demand on academic knowledge, (b) instructions that were presented visually rather than orally, and (c) answers that do not require verbal expression would yield similar mean scores across demographic variables. More specifically, we examined race, ethnicity, gender, and parental education level (PEL) differences on a prepublication form of the *Naglieri General Ability Tests* (Naglieri, Brulles & Lansdowne, 2021), which is comprised of three separate tests: the Verbal (Naglieri–V; Naglieri & Brulles, 2021), Quantitative (Naglieri–Q; Naglieri & Lansdowne, 2021), and Nonverbal (Naglieri–NV; Naglieri, 2021).

Research Question 1: Do students who closely represent the U.S. population on a variety of demographic variables earn similar mean scores by race on each of the Verbal, Quantitative, and Nonverbal Naglieri General Ability Tests?

Research Question 2: Do students who closely represent the U.S. population on a variety of demographic variables earn similar mean scores by ethnicity on each of the Verbal, Quantitative, and Nonverbal Naglieri General Ability Tests?

Research Question 3: Do students who closely represent the U.S. population on a variety of demographic variables earn similar mean scores by gender on each of the Verbal, Quantitative, and Nonverbal Naglieri General Ability Tests?

Research Question 4: Do students who closely represent the U.S. population on a variety of demographic variables earn similar mean scores by parental educational level on each of the Verbal, Quantitative, and Nonverbal Naglieri General Ability Tests?

Methods

Participants

The participants in this study were 7,234 children and adolescents from 12 states in each of the four regions of the US enrolled in Pre-K through Grade 12. Data collection was conducted by an independent research firm which administered the tests in centers on computers. To ensure that administration procedures are strictly followed, examiners employed by the professional data collection company were explicitly trained in proper data collection techniques and protocols. Additionally, each examiner attended a two-day training provided by the researchers to ensure proper administration of the tests. The firm conducted the three separate research projects involving public and private school students from 12 different sites in the four regions of the U.S. in late 2017 and 2018. Each of the three separate groups received one of the following: Naglieri–V, Naglieri–NV, or Naglieri–Q tests. Consent to participate was given by the parent of every participant under the age of 18 and those 18 and older provided consent for themselves. Parents provided basic demographic information for the student in the study, as well as reporting the primary parent's highest level of education. All participants received monetary compensation for their participation. The samples for the Naglieri General Ability Test: Verbal (N = 2,482), Nonverbal (N = 2,383), and Quantitative (N = 2,369) were selected to match the recent U.S. census proportions (U.S. Bureau of the Census, 2016) within two percentage points by gender, race, ethnicity, region, and PEL; see Table 1A through Table 1H for detailed descriptions of all three samples by grade. One of the goals of test development was to create Pre-Kindergarten versions of the Verbal and Quantitative tests, however, during beta testing with Pre-Kindergarten children, it became apparent that they could not understand the instructions clearly and could not complete the Verbal and Quantitative tests. It was not apparent exactly why these children had difficulty, but the likely reason was limited knowledge of verbal concepts and numbers. For this reason, Pre-Kindergarten students were not included in samples for the Verbal and Quantitative tests.

Table 1. a. Demographic Characteristics of the Samples: Pre-Kindergarten.

Demographic Characteristic			Nonverbal	
		Census %	*N*	%
Gender	Male	50.0	155	47.8
	Female	50.0	165	50.9
Race/Ethnic Group	Asian	4.9	17	5.2
	Black	13.6	41	12.7
	Hispanic	24.5	75	23.1
	White	51.4	167	51.5
	Other	5.8	20	6.2
Region	Northeast	16.1	51	15.7
	Midwest	21.2	57	17.6
	South	38.5	119	36.7
	West	24.2	93	28.7
Parental Education Level (PEL)	No high school diploma	15.2	45	13.9
	High school graduate	27.3	81	25.0
	Some college or associate's degree	28.7	90	27.8
	Bachelor's degree	17.9	50	15.4
	Graduate or professional degree	10.8	35	10.8
	Age in years *M (SD)*		4.7 (0.8)	
	Total		324	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students. PEL information was unavailable for 23 students. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. b. Demographic Characteristics of the Samples: Kindergarten.

Demographic Characteristic			Verbal		Nonverbal		Quantitative	
		Census %	*N*	%	*N*	%	*N*	%
Gender	Male	50.0	177	50.0	148	48.1	166	48.5
	Female	50.0	177	50.0	156	50.6	176	51.5
Race/Ethnic Group	Asian	4.9	19	5.4	16	5.2	19	5.6
	Black	13.6	43	12.1	38	12.3	46	13.5
	Hispanic	24.5	84	23.7	81	26.3	83	24.3
	White	51.4	188	53.1	154	50.0	175	51.2
	Other	5.8	20	5.6	15	4.9	19	5.6
Region	Northeast	16.1	58	16.4	50	16.2	55	16.1
	Midwest	21.2	71	20.1	62	20.1	73	21.3
	South	38.5	141	39.8	113	36.7	130	38.0
	West	24.2	84	23.7	79	25.6	84	24.6
Parental Education Level (PEL)	No high school diploma	15.2	56	15.8	45	14.6	52	15.2
	High school graduate	27.3	91	25.7	79	25.6	97	28.4
	Some college or associate's degree	28.7	103	29.1	88	28.6	96	28.1
	Bachelor's degree	17.9	66	18.6	55	17.9	64	18.7
	Graduate or professional degree	10.8	38	10.7	33	10.7	33	9.6
Age in years *M (SD)*			5.2 (0.6)		5.8 (0.7)		5.6 (0.6)	
Total			354		308		342	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 8 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. c. Demographic Characteristics of the Samples: Grade 1.

Demographic Characteristic			Verbal		Nonverbal		Quantitative	
		Census %	*N*	%	*N*	%	*N*	%
Gender	Male	50.0	174	50.6	153	51.0	166	50.2
	Female	50.0	170	49.4	143	47.7	165	49.8
Race/Ethnic Group	Asian	4.9	20	5.8	17	5.7	16	4.8
	Black	13.6	47	13.7	38	12.7	41	12.4
	Hispanic	24.5	81	23.5	73	24.3	82	24.8
	White	51.4	174	50.6	150	50.0	172	52.0
	Other	5.8	22	6.4	18	6.0	20	6.0
Region	Northeast	16.1	57	16.6	53	17.7	53	16.0
	Midwest	21.2	69	20.1	54	18.0	70	21.1
	South	38.5	136	39.5	112	37.3	128	38.7
	West	24.2	82	23.8	77	25.7	80	24.2
Parental Education Level (PEL)	No high school diploma	15.2	53	15.4	44	14.7	51	15.4
	High school graduate	27.3	92	26.7	80	26.7	89	26.9
	Some college or associate's degree	28.7	100	29.1	80	26.7	97	29.3
	Bachelor's degree	17.9	65	18.9	53	17.7	60	18.1
	Graduate or professional degree	10.8	34	9.9	35	11.7	34	10.3
Age in years *M (SD)*			6.3 (0.7)		6.8 (0.7)		6.8 (0.7)	
Total			344		300		331	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 8 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. d. Demographic Characteristics of the Samples: Grade 2.

Demographic Characteristic			Verbal		Nonverbal		Quantitative	
		Census %	*N*	%	*N*	%	*N*	%
Gender	Male	50.0	172	49.7	139	49.6	163	49.1
	Female	50.0	174	50.3	137	48.9	169	50.9
Race/Ethnic Group	Asian	4.9	19	5.5	15	5.4	17	5.1
	Black	13.6	45	13.0	40	14.3	42	12.7
	Hispanic	24.5	85	24.6	67	23.9	81	24.4
	White	51.4	177	51.2	140	50.0	172	51.8
	Other	5.8	20	5.8	14	5.0	20	6.0
Region	Northeast	16.1	57	16.5	46	16.4	60	18.1
	Midwest	21.2	74	21.4	51	18.2	68	20.5
	South	38.5	131	37.9	104	37.1	124	37.3
	West	24.2	84	24.3	75	26.8	80	24.1
Parental Education Level (PEL)	No high school diploma	15.2	55	15.9	41	14.6	51	15.4
	High school graduate	27.3	91	26.3	73	26.1	91	27.4
	Some college or associate's degree	28.7	100	28.9	84	30.0	98	29.5
	Bachelor's degree	17.9	63	18.2	48	17.1	57	17.2
	Graduate or professional degree	10.8	37	10.7	28	10.0	35	10.5
Age in years *M (SD)*			7.3 (0.6)		7.8 (0.7)		7.8 (0.7)	
Total			346		280		332	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 6 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. e Demographic Characteristics of the Samples: Grade 3–4.

Demographic Characteristic			Verbal		Nonverbal		Quantitative	
		Census %	*N*	%	*N*	%	*N*	%
Gender	Male	50.0	181	49.6	151	51.2	174	49.9
	Female	50.0	184	50.4	140	47.5	175	50.1
Race/Ethnic Group	Asian	4.9	20	5.5	11	3.7	18	5.2
	Black	13.6	48	13.2	42	14.2	43	12.3
	Hispanic	24.5	85	23.3	73	24.7	86	24.6
	White	51.4	191	52.3	150	50.8	182	52.1
	Other	5.8	21	5.8	15	5.1	20	5.7
Region	Northeast	16.1	55	15.1	52	17.6	64	18.3
	Midwest	21.2	75	20.5	57	19.3	71	20.3
	South	38.5	145	39.7	107	36.3	133	38.1
	West	24.2	90	24.7	75	25.4	81	23.2
Parental Education Level (PEL)	No high school diploma	15.2	55	15.1	38	12.9	55	15.8
	High school graduate	27.3	95	26.0	83	28.1	94	26.9
	Some college or associate's degree	28.7	108	29.6	85	28.8	98	28.1
	Bachelor's degree	17.9	65	17.8	53	18.0	63	18.1
	Graduate or professional degree	10.8	42	11.5	27	9.2	39	11.2
Age in years *M (SD)*			8.7 (0.9)		9.3 (0.8)		9.4 (0.9)	
Total			365		295		349	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 9 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. f Demographic Characteristics of the Samples: Grade 5–6.

Demographic Characteristic		Census %	Verbal		Nonverbal		Quantitative	
			N	%	*N*	%	*N*	%
Gender	Male	50.0	175	50.0	145	49.5	163	47.4
	Female	50.0	175	50.0	144	49.1	181	52.6
Race/Ethnic Group	Asian	4.9	19	5.4	11	3.8	20	5.8
	Black	13.6	46	13.1	40	13.7	47	13.7
	Hispanic	24.5	83	23.7	73	24.9	79	23.0
	White	51.4	183	52.3	149	50.9	176	51.2
	Other	5.8	19	5.4	16	5.5	22	6.4
Region	Northeast	16.1	56	16.0	53	18.1	60	17.4
	Midwest	21.2	73	20.9	51	17.4	70	20.3
	South	38.5	138	39.4	110	37.5	132	38.4
	West	24.2	83	23.7	75	25.6	82	23.8
Parental Education Level (PEL)	No high school diploma	15.2	57	16.3	42	14.3	52	15.1
	High school graduate	27.3	92	26.3	77	26.3	101	29.4
	Some college or associate's degree	28.7	101	28.9	87	29.7	95	27.6
	Bachelor's degree	17.9	64	18.3	46	15.7	60	17.4
	Graduate or professional degree	10.8	36	10.3	34	11.6	36	10.5
Age in years *M (SD)*			10.8 (0.8)		11.1 (1.0)		11.3 (0.9)	
Total			350		293		344	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 7 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. g Demographic Characteristics of the Samples: Grade 7–9.

Demographic Characteristic		Census %	Verbal		Nonverbal		Quantitative	
			N	%	*N*	%	*N*	%
Gender	Male	50.0	180	49.0	141	48.0	166	49.1
	Female	50.0	187	51.0	149	50.7	172	50.9
Race/Ethnic Group	Asian	4.9	19	5.2	10	3.4	16	4.7
	Black	13.6	47	12.8	40	13.6	44	13.0
	Hispanic	24.5	85	23.2	71	24.1	79	23.4
	White	51.4	196	53.4	152	51.7	181	53.6
	Other	5.8	20	5.4	16	5.4	18	5.3
Region	Northeast	16.1	60	16.3	57	19.4	61	18.0
	Midwest	21.2	74	20.2	54	18.4	69	20.4
	South	38.5	145	39.5	113	38.4	133	39.3
	West	24.2	88	24.0	66	22.4	75	22.2
Parental Education Level (PEL)	No high school diploma	15.2	56	15.3	43	14.6	49	14.5
	High school graduate	27.3	100	27.2	76	25.9	89	26.3
	Some college or associate's degree	28.7	106	28.9	85	28.9	101	29.9
	Bachelor's degree	17.9	62	16.9	52	17.7	56	16.6
	Graduate or professional degree	10.8	43	11.7	31	10.5	43	12.7
Age in years *M (SD)*			13.2 (1.1)		13.7 (1.0)		13.8 (1.0)	
Total			367		294		338	

Note. Gender, Race/Ethnicity, and Region information was unavailable for 4 students in the Nonverbal sample. PEL information was unavailable for 7 students in the Nonverbal sample. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Table 1. h Demographic Characteristics of the Samples: Grade 10–12.

Demographic Characteristic			Verbal		Nonverbal		Quantitative	
		Census %	*N*	%	*N*	%	*N*	%
Gender	Male	50.0	177	49.7	139	48.1	161	48.3
	Female	50.0	179	50.3	146	50.5	172	51.7
Race/Ethnic Group	Asian	4.9	19	5.3	13	4.5	20	6.0
	Black	13.6	43	12.1	33	11.4	46	13.8
	Hispanic	24.5	85	23.9	78	27.0	75	22.5
	White	51.4	188	52.8	141	48.8	173	52.0
	Other	5.8	21	5.9	20	6.9	19	5.7
Region	Northeast	16.1	63	17.7	47	16.3	59	17.7
	Midwest	21.2	77	21.6	49	17.0	68	20.4
	South	38.5	133	37.4	111	38.4	132	39.6
	West	24.2	83	23.3	78	27.0	74	22.2
Parental Education Level (PEL)	No high school diploma	15.2	52	14.6	41	14.2	50	15.0
	High school graduate	27.3	93	26.1	79	27.3	83	24.9
	Some college or associate's degree	28.7	100	28.1	82	28.4	100	30.0
	Bachelor's degree	17.9	65	18.3	46	15.9	62	18.6
	Graduate or professional degree	10.8	46	12.9	32	11.1	38	11.4
Age in years *M (SD)*			16.3 (1.3)		16.9 (1.4)		16.9 (1.2)	
Total			356		289		333	

Note. Census values reflect proportions derived from the 2016 American Community Survey (US Census Bureau, 2016).

Administration

The instructions and user interface for the online tests were developed to reduce the demand for verbal directions. In each test, an animated instructional video was developed to show the student how to complete the task. The video portrays a child sitting at a computer and attempting to complete a sample test item. The child's thinking process is illustrated by a thought bubble that shows how the child arrives at the correct answer. The video also illustrates how the child can select a response, change their response, and navigate to the next item. The administrator does not need to provide any verbal directions throughout this video–but they are given the opportunity to provide additional help as needed so that the student understands the requirements of the task.

The practice items that follow the instructions were also designed to require minimal verbal assistance. They were programmed to be interactive so that students are able to learn the correct answer by selecting any response. When the incorrect response is selected, the forward arrow is disabled, and the correct response is visually highlighted. The forward arrow is only enabled when the correct response is selected, thus allowing students to learn the correct answer without any verbal input from the administrator.

The administration procedure was identical for all three tests. All tests were administered to students on a computer, and their responses were recorded automatically. All testing sessions included an administrator who monitored the students during testing. Administrators ensured that all students knew how to use the computers and provided assistance to any students who required it. Pre-Kindergarten students were tested in a one-on-one setting with the administrator sitting next to them, while Kindergarten students were tested in small groups of two to three students at a time. Testing for students in Grades 1–2 and Grades 3–12 occurred in larger group sizes of four to five students at a time and up to 10 students at a time, respectively. For all group sessions, the administrator provided minimal verbal directions from the front of the room. In each testing session, the student(s) watched the animated instructional video and then completed three interactive practice items. The students were then prompted to begin the test and were given 45 minutes to complete the test items. They were allowed to skip items, as well as navigate backwards to earlier items.

Instruments

The Naglieri–V (Naglieri & Brulles, 2021) test was developed as a measure of general ability using questions which require a student to recognize a verbal concept when that concept is represented by pictures. All items consist of six images displayed in a 2x3 grid. The student must determine which verbal concept is shared by five out of six pictures and then select the picture that does not represent the concept (e.g., five different types of fruit, and a wicker basket). These items test a student's ability to reason with verbal concepts in any language – that is, categorize ideas or things based on their name, function, and/or characteristics. This task is based on Luria's (1982) concept of the superfluous fourth which requires a person to determine which word is different from several others that signify similar concepts (e.g., living things). Item creation was guided by the inclusion of various combinations of four aspects of language denoting objects, attributes, actions, or relationships. This approach was used instead of traditional vocabulary or word analogies tests that demand knowledge typically associated with educational opportunity in and outside of school.

The test items are organized into seven levels (K, Grade 1, Grade 2, Grade 3-4, Grade 5-6, Grade 7-9, and Grade 10-12), each containing 50 items selected to be most appropriate for children at the grade or grades for which that level is intended. Each level contains some items that are unique to that level and some that are shared across adjacent grade levels. The internal consistency coefficients (as measured by coefficient alpha; Cronbach, 1951) for the Verbal test by level ranged from .83 to .92 for the present sample. Concerning measurement bias, fewer than 6% of all items across all forms and all three tests, in total, were flagged with respect to a logistic regression differential item functioning (DIF) analysis for gender, race, and PEL. In many instances, the detected DIF is a very small effect size or only occurs at a certain theta level and therefore does not confer cause for concern. The absence of meaningful DIF warrants the investigation and interpretation of mean differences in the present study. As a detailed account of measurement bias is beyond the scope of this paper, interested readers are encouraged to review the technical manual where comprehensive DIF and differential test functioning (DTF) analyses are included (Naglieri et al., 2021).

The Naglieri–Nonverbal (Naglieri, 2021) test was developed to measure general ability using questions which require a student to carefully examine elements presented in a pictorial or matrix format (e.g., 2x2, 2x3, or 3x3) where shapes, sequences, spatial orientations, movement patterns, and other distinguishing characteristics must be analyzed to solve the question. That is, the student must examine the visual components in the matrix and figure out a rule that explains the relationship among those components. That rule must then be applied to arrive at the answer (one of five options presented below the matrix). This test is similar to other progressive matrices tests such as the *Raven's Progressive Matrices* (Raven & Raven, 2003) and the *Naglieri Nonverbal Ability Test – Third Edition* (Naglieri, 2016). The Naglieri–Nonverbal is conceptually very similar to the various editions of the NNAT in that all these tests were designed to measure general ability using diagrams. The new version includes original and new formats used to generate the items across the range of difficulty required.

The test items are organized into eight levels (Pre-Kindergarten, Kindergarten, Grade 1, Grade 2, Grade 3-4, Grade 5-6, Grade 7-9, and Grade 10-12), each containing 55 items selected to be most appropriate for children at the grade or grades for which that level is intended. Each level contains some items that are unique to that level and some that are shared across adjacent grade levels. The internal consistency reliability coefficients for the Nonverbal test by level ranged from .85 to .92 across grade levels for the present sample. A trivial number of items were flagged with respect to a DIF analysis for gender, race, and PEL. A detailed account of measurement bias is beyond the scope of this paper. Interested readers are encouraged to review the technical manual where comprehensive DIF and DTF analyses are included (Naglieri et al., 2021).

The Naglieri–Q (Naglieri & Lansdowne, 2021) test was developed to measure general ability using questions that require a student to closely examine the relationships among the numbers and/or symbols using basic math concepts (e.g., addition, subtraction, multiplication, division). The student must look at the components of the question and figure out a quantitative rule that explains the relationship among those components. That rule must then be applied to arrive at the answer, indicated by selecting one of five options presented below the question. The items were written to demand minimal mathematical knowledge instead of math word problems or math reasoning tasks with verbal answers. The test items are organized into seven levels (K, Grade 1, Grade 2, Grade 3-4, Grade 5-6, Grade 7-9, and Grade 10-12),

each containing 50 items designed to be appropriate for children at the grade or grades for which that level is intended. The internal consistency reliability coefficients for the Quantitative test by level ranged from .87 to .93 across grade levels for the present sample. As found with the Verbal and Nonverbal versions, a trivial number of items were flagged from a DIF analysis for gender, race, and PEL. See the technical manual for a comprehensive analysis of DIF and DTF (Naglieri et al., 2021).

The three tests were constructed to measure general ability using three different types of items but we do not assume that three different abilities are being measured; rather, these tests could be used to measure what is commonly described as Spearman's g. That is, the consistent finding of positive correlations between tests with varying content which has recently been explained on the basis of overlapping cognitive processes that represent the general factor of intelligence, g, as a formative construct (Kovacs & Conway, 2019).Recent research of widely used intelligence tests has consistently suggested that the most valid score on the (a) Wechsler Intelligence Scale for Children – Fifth Edition (Canivez et al., 2017), (b) Stanford-Binet Fifth Edition (Canivez, 2008), (c) Differential Abilities Scales (Canivez & McGill, 2016), and (d) Woodcock-Johnson Fourth Edition (Dombrowski, McGill & Canivez, 2017) is the total score that estimates *g* (i.e., general ability). These researchers found that the scores which represent the factor-based scales included in these tests do not have enough specific variance to be considered interpretable. Similarly, a reanalysis of Carroll's survey of factor-analytic studies (Carroll, 1993) by Benson et al. (2018) concluded that nearly all the specified abilities presented by Carroll "have little-to-no interpretive relevance above and beyond that of general intelligence (p. 1028)." Using similar factorial methodologies, Naglieri, Brulles, and Lansdowne (2021) provided evidence that the three Naglieri General Ability Tests are primarily an assessment of general ability and are consistent with scientific models of intelligence (e.g., Benson et al., 2018, Canivez & McGill, 2016).

Data Analysis

Data were analyzed using R (R Core Team, 2013) and R Studio (R Studio Team, 2019), primarily with the package *rstatix* (Kassambara, 2021). First, data were screened and cleaned for valid responses (e.g., age and date of birth correspond, omitted items, or items not reached during the time limit were coded as incorrect responses). Items were scored and the total number of correct responses was summed to create a total score for each test (note that the maximum total score for Nonverbal was 55 items correct, and for Verbal and Quantitative, it was 50 items correct). Raw scores (which could range from 0 to 55 points for Nonverbal and 0 to 50 for Verbal and Quantitative) were analyzed and found to be normally distributed.

Analyses were conducted to examine the influence of each demographic variable as a meaningful predictor of Naglieri General Ability Test scores. Relations between demographic variables (gender, PEL, and race/ethnicity) and total scores (i.e., the sum of correct responses) were investigated with independent t-tests (for gender, as it was coded as binary [male, female]) and ANOVAs (for PEL and race/ethnicity) to inspect observed mean differences between groups. For these two variables of interest, U.S. census categorization was used; PEL is presented and analyzed with five levels, and for race/ethnicity, race in this study is defined as non-Hispanic White youth and non-Hispanic Black youth, and ethnicity (for the purposes of this study, and in accordance with the U.S. census) refers to Hispanic youth. Note that sample sizes for Asian students were too small to permit further investigation due to limited power. Statistical significance was corrected with Bonferroni *p*-value adjustment to control the family-wise error rate of multiple comparisons (e.g., alpha of .05 divided by seven grade-level comparisons = adjusted *p*-value of .007).

Results

Overall, main effects of race, ethnicity, gender, and PEL were explored, and minimal differences were observed. Results of analyses of group differences by race (White vs. Black) and ethnicity (White vs. Hispanic) are presented for the Verbal, Nonverbal, and Quantitative tests in Tables 2a, 2b, and 2c, respectively. Using the adjusted *p*-value to determine significance, the differences between racial/ethnic groups were not statistically significant for all grade-levels for all three tests. The size of the differences were negligible (median $\eta^2 = 0.01$), and pairwise differences between White and Black, and White and Hispanic, groups were also negligible (median Cohen's $d = 0.02$). These results address the first two research hypotheses, finding no evidence for significant differences between the groups in this study.

Table 2.a Race & Ethnicity Differences: Naglieri-Verbal.

Grade-Form	Group	*N*	*M*	*SD*	*F*	*df*	*p*	η^2
Kindergarten	Black	43	32.05	10.90	2.32	2, 312	.100	0.01
	Hispanic	84	34.33	11.78				
	White	188	31.21	10.72				
Grade 1	Black	47	27.96	7.47	1.77	2, 299	.172	0.01
	Hispanic	81	31.20	10.87				
	White	174	30.39	9.39				
Grade 2	Black	45	26.04	9.44	3.34	2, 304	.037	0.02
	Hispanic	85	30.48	10.98				
	White	177	29.86	9.31				
Grade 3-4	Black	48	29.52	8.70	3.48	2, 321	.032	0.02
	Hispanic	85	32.80	7.66				
	White	191	32.68	7.55				
Grade 5-6	Black	46	32.98	4.72	0.31	2, 309	.737	0.00
	Hispanic	83	33.83	7.07				
	White	183	33.23	7.11				
Grade 7-9	Black	47	29.72	6.86	1.32	2, 325	.270	0.01
	Hispanic	85	31.53	8.07				
	White	196	31.50	6.49				
Grade 10-12	Black	43	29.74	8.68	3.97	2, 313	.020	0.02
	Hispanic	85	33.40	6.54				
	White	188	31.22	7.68				

Table 2. b Race & Ethnicity Differences: Naglieri-Nonverbal.

Grade-Form	Group	*N*	*M*	*SD*	*F*	*df*	*p*	η^2
Pre-Kindergarten	Black	41	27.49	12.60	4.65	2, 280	.010	0.03
	Hispanic	75	33.08	10.42				
	White	167	28.81	11.34				
Kindergarten	Black	38	27.63	9.05	0.51	2, 270	.602	0.00
	Hispanic	81	29.43	10.14				
	White	154	29.25	9.47				
Grade 1	Black	38	24.05	8.74	1.28	2, 258	.279	0.01
	Hispanic	73	26.85	7.69				
	White	150	25.91	9.18				
Grade 2	Black	40	26.48	9.53	0.03	2, 244	.970	0.00
	Hispanic	67	26.16	7.20				
	White	140	26.44	8.10				
Grade 3-4	Black	42	26.45	8.59	0.63	2, 262	.536	0.00
	Hispanic	73	26.55	7.38				
	White	150	27.57	7.52				
Grade 5-6	Black	40	28.10	8.07	0.15	2, 259	.864	0.00
	Hispanic	73	28.41	7.36				
	White	149	27.83	7.64				
Grade 7-9	Black	40	28.40	9.10	0.17	2, 260	.841	0.00
	Hispanic	71	28.56	8.97				
	White	152	27.83	9.49				
Grade 10-12	Black	33	24.94	6.49	3.32	2, 249	.038	0.03
	Hispanic	78	27.86	9.57				
	White	141	29.50	9.89				

Table 2. c Race & Ethnicity Differences: Naglieri-Quantitative.

Grade-Form	Group	*N*	*M*	*SD*	*F*	*df*	*p*	η^2
Kindergarten	Black	46	29.65	11.69	4.17	2, 301	.016	0.03
	Hispanic	83	34.86	10.16				
	White	175	31.86	10.12				
Grade 1	Black	41	27.00	9.66	5.51	2, 292	.004	0.04
	Hispanic	82	33.60	10.24				
	White	172	31.00	10.71				
Grade 2	Black	42	24.76	10.01	4.58	2, 292	.011	0.03
	Hispanic	81	30.35	9.96				
	White	172	27.76	9.81				
Grade 3-4	Black	43	24.28	10.59	2.12	2, 308	.122	0.01
	Hispanic	86	28.07	11.07				
	White	182	26.75	9.07				
Grade 5-6	Black	47	19.26	7.31	3.37	2, 299	.036	0.02
	Hispanic	79	23.28	8.72				
	White	176	22.13	8.67				
Grade 7-9	Black	44	20.20	8.18	3.91	2, 301	.021	0.03
	Hispanic	79	24.08	7.73				
	White	181	23.63	7.91				
Grade 10-12	Black	46	20.28	7.88	1.22	2, 291	.297	0.01
	Hispanic	75	22.08	8.16				
	White	173	22.42	8.38				

Results of the group differences analyses by gender are presented for the Verbal, Nonverbal, and Quantitative tests in Tables 3a, 3b, and 3c, respectively. To summarize, all grade-levels for all three tests revealed a nonsignificant difference between male and female students in terms of raw test scores. Effect sizes were negligible to small, with Cohen's *d*-values ranging from -0.26 to 0.23. The inconsistent direction of the observed differences in scores was also inconsistent (e.g., for some grade-levels, male students scored higher than female students, but that trend reversed for other grade-levels). These results address the third research hypothesis by the absence of a significant difference in scores based on gender groups.

Table 3. a Gender Differences: Naglieri-Verbal.

Grade-Form	**Gender**	***N***	***M***	***SD***	***t***	***df***	***p***	**Cohen's *d***
Kindergarten	Male	177	32.88	10.64	0.52	351	.605	0.05
	Female	177	32.28	11.14				
Grade 1	Male	174	30.21	9.88	-0.13	341	.901	-0.01
	Female	170	30.34	9.15				
Grade 2	Male	172	29.33	9.47	-1.06	343	.291	-0.11
	Female	174	30.44	10.07				
Grade 3-4	Male	181	31.27	8.33	-2.48	355	.013	-0.26
	Female	184	33.30	7.27				
Grade 5-6	Male	175	33.11	7.13	-1.25	344	.213	-0.13
	Female	175	34.01	6.38				
Grade 7-9	Male	180	30.94	6.54	-1.22	364	.223	-0.13
	Female	187	31.82	7.15				
Grade 10-12	Male	177	31.34	8.13	-1.33	341	.184	-0.14
	Female	179	32.39	6.72				

Table 3. b Gender Differences: Naglieri-Nonverbal.

Grade-Form	Gender	*N*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Pre-Kindergarten	Male	155	30.67	11.81	1.44	311	.150	0.16
	Female	165	28.84	10.78				
Kindergarten	Male	148	28.27	9.40	-1.66	302	.098	-0.19
	Female	156	30.09	9.70				
Grade 1	Male	153	25.93	8.51	-0.12	291	.905	-0.01
	Female	143	26.05	8.80				
Grade 2	Male	139	26.29	8.57	-0.44	274	.658	-0.05
	Female	137	26.74	8.28				
Grade 3-4	Male	151	26.80	7.85	-0.87	289	.386	-0.10
	Female	140	27.59	7.57				
Grade 5-6	Male	145	27.70	7.51	-1.13	287	.261	-0.13
	Female	144	28.71	7.67				
Grade 7-9	Male	141	28.52	9.49	0.64	286	.520	0.08
	Female	149	27.81	9.32				
Grade 10-12	Male	139	29.71	10.15	1.52	278	.129	0.18
	Female	146	27.95	9.34				

Table 3. c Gender Differences: Naglieri-Quantitative .

Grade-Form	Gender	*N*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Kindergarten	Male	166	32.61	10.66	0.05	338	.960	0.01
	Female	176	32.56	10.52				
Grade 1	Male	166	31.21	9.95	-0.25	325	.804	-0.03
	Female	165	31.50	11.03				
Grade 2	Male	163	28.40	10.42	0.10	329	.919	0.01
	Female	169	28.28	10.11				
Grade 3-4	Male	174	28.22	10.21	2.11	346	.036	0.23
	Female	175	25.96	9.85				
Grade 5-6	Male	163	23.15	8.57	1.88	336	.061	0.20
	Female	181	21.44	8.34				
Grade 7-9	Male	166	24.34	8.08	1.93	334	.055	0.21
	Female	172	22.68	7.72				
Grade 10-12	Male	161	22.75	8.48	1.86	320	.064	0.20
	Female	172	21.12	7.49				

Results for the analysis of PEL differences for the Verbal, Nonverbal, and Quantitative tests are presented in Tables 4a, 4b, and 4c, and they present no statistically significant effects between groups. All comparisons yield p-values that exceed the Bonferroni adjusted p-value, and effect sizes are negligible (median $\eta 2 = 0.01$). Scores between different levels of parental education were so close that post-hoc comparisons were not needed, given the nonsignificant omnibus effects observed. These results address the fourth research question, demonstrating an absence of significant differences between student scores based on PEL groups.

Table 4.a Parental Education Level (PEL) Differences: Naglieri-Verbal.

Grade-Form	Group	*N*	*M*	*SD*	*F*	*df*	*p*	η^2
Kindergarten	PEL 1	56	33.84	11.12	0.29	4, 349	.883	0.00
	PEL 2	91	31.89	10.98				
	PEL 3	103	32.46	11.67				
	PEL 4	66	32.45	9.94				
	PEL 5	38	32.95	9.96				
Grade 1	PEL 1	53	30.96	9.01	0.17	4, 339	.955	0.00
	PEL 2	92	30.15	8.74				
	PEL 3	100	29.75	10.28				
	PEL 4	65	30.51	10.46				
	PEL 5	34	30.59	8.42				
Grade 2	PEL 1	55	30.15	10.69	0.43	4, 341	.790	0.00
	PEL 2	91	29.52	9.69				
	PEL 3	100	29.21	9.89				
	PEL 4	63	30.40	9.23				
	PEL 5	37	31.41	9.46				
Grade 3-4	PEL 1	55	30.67	7.77	1.65	4, 360	.161	0.02
	PEL 2	95	31.99	8.47				
	PEL 3	108	32.06	7.99				
	PEL 4	65	33.09	7.11				
	PEL 5	42	34.50	7.07				
Grade 5-6	PEL 1	57	33.86	6.00	0.13	4, 345	.973	0.00
	PEL 2	92	33.59	6.87				
	PEL 3	101	33.19	6.86				
	PEL 4	64	33.81	6.17				
	PEL 5	36	33.61	8.53				
Grade 7-9	PEL 1	56	29.98	8.29	0.96	4, 362	.430	0.01
	PEL 2	100	31.16	7.17				
	PEL 3	106	31.66	6.29				
	PEL 4	62	31.89	6.77				
	PEL 5	43	32.37	5.45				
Grade 10-12	PEL 1	52	31.81	7.66	0.17	4, 351	.954	0.00
	PEL 2	93	31.85	7.68				
	PEL 3	100	32.13	8.07				
	PEL 4	65	31.26	6.79				
	PEL 5	46	32.26	6.53				

Note. PEL 1 = No high school diploma. PEL 2 = High school diploma or equivalent. PEL 3 = Some college or associate's degree. PEL 4 = Bachelor's degree. PEL 5 = Graduate/professional degree.

Discussion

The purpose of this study was to examine group differences in performance across key demographic variables of race, ethnicity, gender, and PEL on an initial version of the *Naglieri General Ability Tests: Verbal, Nonverbal, and Quantitative*. We hypothesized that the way these tests were constructed (i.e., animated instructions, test questions that could be solved regardless of the language used by the examinees, and the elimination of a verbal response to the questions) might reduce traditionally observed differences in student performance across racial/ethnic, gender, and PEL groups. Additionally, the test development goals were designed to maximize the likelihood that examinees could solve verbal, nonverbal, and quantitative test questions and measure general ability in a manner that was minimally influenced by academic knowledge.

The present findings suggest minimal nonsignificant differences in scores were found for White, Black, and Hispanic students across all forms of the three tests. These results are in contrast to previous studies that have shown that Black and Hispanic students consistently score lower than White students

on the most widely used ability tests (Brulles et al., 2022; Brody, 1992; Helms, 1992; Miller-Jones, 1989; Naglieri & Otero, 2017; Pfeiffer, 2012; Sattler, 1988; Skiba et al., 2002; Wasserman & Becker, 2000). Our results are consistent with researchers who have suggested that nonverbal tests of general ability have value when assessing diverse populations because of their reduced emphasis on language on the part of both the examiner and the student and minimal academic demands (Brulles et. al., 2022; Athanasiou, 2000; McCallum et al., 2001, Naglieri & Ford, 2005).

The negligible mean score differences between males and females found on the Naglieri-Verbal test are consistent with previous studies which found similar scores on verbal reasoning but higher scores for females than males in reading and writing (Hyde & Linn, 1988; Lohman & Lakin, 2009; Strand et al., 2006). For example, an analysis of multiple U.S. national datasets for students assessed across multiple age groups on a variety of tests indicated no gender differences on verbal reasoning (d = 0.05) and a large difference in writing (d = 0.60) in favor of females (Cole, 1997). Given that the present study utilized a pictorial item format that allowed verbal reasoning in any language, thus eliminating the demand for reading and writing skills, the present study's negligible gender differences make sense.

The negligible differences in mean scores between males and females found on the Naglieri-Nonverbal test are similar to findings reported on tests such as the *Raven's Progressive Matrices* (Raven & Raven, 2003) and the *Naglieri Nonverbal Ability Test* (Naglieri, 2016), which also show small gender differences (Court, 1983; Lynn & Irwing, 2004). These findings are consistent with Lynn and Irwing (2004) who found that there was no gender difference in mean scores on the Raven's Progressive Matrices for the age range of 6- to 14- years-old. The negligible gender differences reported here are also consistent with Rojahn and Naglieri (2006), who found trivial gender differences in mean scores using the *Naglieri Nonverbal Ability Test* (Naglieri, 1997) standardization sample of children and adolescents in grades K-12 (aged 6 to 17 years; N = 79,780), which was representative of the U.S. population on several critical demographic variables.

Negligible differences in mean scores between males and females were also found on the Naglieri-Quantitative test. Mathematical skills have long been described as a male domain (Fennema & Sherman, 1977; Hyde et al., 1990a, Nosek et al., 2009). Such beliefs exist among children and adolescents, parents, and teachers (Steele, 2003; Furnham et al., 2002; Frome & Eccles, 1998; Li, 1999; Helwig et al., 2001). Although male dominance in math have been reported on math achievement tests (Halpern, 1992), no advantage was observed with the present Quantitative test, perhaps because of the reduced knowledge demands. These findings are consistent with previous research on other quantitative measures (Hyde et al., 1990b; Hedges and Nowell, 1995; Strand et al., 2006; Mills et al., 1993).

Significant differences between PEL groups, used here as a single indicator and proxy for socioeconomic status (SES), were not found on any of the three *Naglieri General Ability Tests*. The lack of significant differences is in contrast to findings for tests that demand academic knowledge. Tyler-Wood and Carri (1993) found that the gap between low SES and average SES students' test scores was highest on the verbal subscales requiring English knowledge on many tests of ability. Researchers typically find lower performance on both achievement and intelligence tests that are loaded with language demands for children with lower SES (Olszewski-Kubilius & Corwith, 2018). Similarly, Warne et al. (2013) found that much of the disparity in the identification and participation of students of color and low-income students in gifted programs in the state of Utah was due to these students' lower academic achievement scores. Ryan and French (1976) found that large differences existed in students from low, middle, and high SES groups on achievement. Portes and MacLeod (1996) found that SES had a strong effect on Stanford Achievement Test scores, even after controlling for several other individual variables. The current findings suggest that the reduction in the academic knowledge required by the Naglieri General Ability Tests may be associated with minimized differences across students with varying levels of SES.

The very small differences across race, ethnicity, gender, and PEL on all three measures of general ability which contain different content (Verbal, Quantitative, and Nonverbal) illustrate that when direct attempts are made to minimize the inherent influences resulting from language and academic knowledge in administration practices and test content, the possibility of a more equitable way of assessing students may be achieved. Addressing the language and knowledge demands in the items as well as the instructions and student response format appears to have helped minimize differences in student performance across groups with respect to race/ethnicity, gender, and PEL. Careful consideration of the language and

knowledge in test content, student responses, and the instructions for administration in the tools used for gifted identification may play a role in equitable assessment. The remaining critical issue is how this different method of identification might be paired with the delivery of a differentiated curriculum that could address the academic needs of students who are smart but not yet knowledgeable (see Brulles et al., 2022).

Limitations

This study, like most, has limitations that should be considered. First, although the characteristics of the three samples were similar to the U.S. population and the groups were large, the three tests were not administered to the same students. This administration procedure limited analysis of the relationships among the three tests, which could provide additional understanding of the validity of the measures. Second, because each test had several forms, the samples could not be combined across age groups, and results had to be presented by grade-level. Consequently, the samples used to examine differences by level were sometimes small, which could potentially reduce the power to detect statistically significant effects. Third, the relationship between these three measures of general ability and academic achievement should be examined to further assess their validity. Given these limitations, we suggest that it appears the Verbal, Nonverbal, and Quantitative tests used in this investigation may hold promise as measures of general ability that are minimally related to race, ethnicity, gender, and PEL. These findings should be supplemented with additional research on the complete psychometric qualities of the final published version of these tests for the normative sample, once published, as well as practical issues such as identification hit rates for high functioning students and direct comparisons of standard scores from other tests of general ability.

Institutional Review Board Statement:

This study of the validity of three general ability tests was jointly conducted by the tests' publisher, Multi-Health Systems (MHS) and a professional research company. MHS has its own internal review board which meticulously examined the research protocol before any data was collected. This board consists of scientists and subject matter experts with considerable knowledge and experience regarding experiments with human subjects and many years of research experience both in institutional and commercial settings. The research protocols were all examined and conducted with the rights and welfare of the human subjects protected. This included, but was not limited to, obtained consent before beginning the study, briefing and debriefing the study participants, and no use of deception. MHS designed and approved the study and the professional research company which has been conducting research for 35 years collected the data using individuals specifically trained in scientific research protocols and data collection procedures.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written consent to participate was given by the parent of every participant under the age of 18 and those 18 and older provided consent for themselves. Parents provided basic demographic information for the student in the study, as well as reporting the primary parent's highest level of education. All participants received monetary compensation for their participation.

Data Availability Statement: Data is not available due to privacy restrictions.

Conflicts of Interest: Multi-Health Systems Inc. is the publisher of the *Naglieri General Ability Tests* and employer of some of the authors of this paper. Naglieri is the author of the tests studied.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied measurement in education*, *14*(3), 219-234.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.

Athanasiou, M. (2000). Current nonverbal assessment instruments: A comparison of psychometric integrity and test fairness. *Journal of Psychoeducational Assessment, 18*(3), 211–229. https://doi.org/10.1177/073428290001800302

Benson, N. F., Beaujean, A. A., McGill, R. J, & Dombrowski, S. C. (2018). Revisiting Carroll's survey of factor-analytic studies: Implications for the clinical assessment of intelligence. *Psychological Assessment, 30*(8), 1028–1038.

Borland, J. H. (2009). Myth 2: The gifted constitute 3% to 5% of the population. Moreover, giftedness equals high IQ, which is a stable measure of aptitude. *Gifted Child Quarterly, 53*, 236–238. https://doi.org/10.1177/0016986209346825

Brody, N. (1992). *Intelligence.* Academic Press.

Brulles, D., Lansdowne, K. & Naglieri, J. A. (2022). *Understanding and Using the Naglieri General Ability Tests: A Call to Equity in Gifted Education.* Free Spirit Publishing.

Brulles, D. & Naglieri, J. A. (2022). A social justice approach to developing equity and diversity in gifted programming. In Nyberg, J. & Manzone, J. (Eds.) *Creating Equitable Services for the Gifted: Protocols for Identification, Implementation, and Evaluation.* https://www.igi-global.com/chapter/a-social-justice-approach-to-developing-equity-and-diversity-in-gifted-programming/290595

Callahan, C. M. (2005). Identifying gifted students from underrepresented populations. *Theory into Practice, 44*, 98-104. http://www.jstor.org/stable/3497028

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29,* 458-472. http://dx.doi.org/10.1037/pas0000358

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales–Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment, 28,* 1475-1488. http://dx.doi.org/10.1037/pas0000279

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. *School Psychology Quarterly, 23,* 533–541. https://doi.org/10.1037/a0012884.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.

Castellano, J. A., & Frazier, A. D. (2010). *Special populations in gifted education: Understanding our most able students from diverse backgrounds*. Prufrock Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cole, N. S. (1997). *ETS gender study: How females and males perform in educational settings.* Education Testing Service.

Court, J. H. (1983). Sex differences in performance on Raven's Progressive Matrices: A review. *Alberta Journal of Educational Research, 29*(1), 54–74.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–335. https://doi.org/10.1007/BF02310555

De Brey, C., Snyder, T. D., Zhang, A., & Dillow, S. A. (2021). *Digest of Education Statistics 2019* (NCES 2021-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Dickens, W., & Flynn, J. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17*(10), 913–920. https://doi.org/10.1111%2Fj.1467-9280.2006.01802.x

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Cognitive at school age. *Psychological Assessment, 29,* 394-407. http://dx.doi.org/10.1037/pas0000350

Faul, F., Erdfelder, E., Lang, A., & Buchner, A (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 39, 175-191.

Fennema, E. & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal, 14*(1), 51-71. https://doi.org/10.3102/00028312014001051

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education, 32*, 4–14. https://doi.org/10.1177/002246699803200102

Ford, D. Y. (2013). *Recruiting and retaining culturally different students in gifted education*. Prufrock Press.

Frasier, M. M., Garcia, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students*. University of Connecticut, National Research Center on the Gifted and Talented.

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74*(2), 435–452. https://doi.org/10.1037/0022-3514.74.2.435

Furnham, A., Reeves, E. & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *Journal of Genetic Psychology, 163*, 24–39. https://doi.org/10.1080/00221320209597966

Gentry, M., Gray, A., Whiting, G., Maeda, Y, & Pereira, N (2019). *Gifted education in the United States: Laws, access, equity and missingness across the country by locale, Title I school status, and race. https://www.education.purdue,edu/geri/new-publications/gifted-education-in-the-united-states/*

Gill, C., Moorer-Cook, L., Armstrong, E. S., & Gill, K. (2012). The Ability to Follow Verbal Directions: Identifying Skill Levels and Measuring Progress. *Canadian Journal of Speech-Language Pathology & Audiology*, *36*(3).

Halpern, D. F. (1992). *Sex differences in cognitive abilities*. Lawrence Erlbaum.

Harcourt Educational Measurement. (1996). *Stanford Achievement Test Series, Ninth Ed*. Harcourt Education Measurement.

Harris, B., Rapp, K., Martínez, R., & Plucker, J. (2007). Identifying English language learners for gifted and talented programs: Current practices and recommendations for improvement. *Roeper Review, 30,* 26–29. https://doi.org/10.1080/02783193.2007.11869221

Hedges, L. V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45. https://www.jstor.org/stable/2889145

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*(9), 1083–1101. https://doi.org/10.1037/0003-066X.47.9.1083

Helwig, R., Anderson, L., & Tindal, G. (2001). Influence of elementary student gender on teachers' perceptions of mathematics achievement. *The Journal of Educational Research, 95*(2), 93–102. https://doi.org/10.1080/00220670109596577

Hodges, J., Tay, J., Maeda, Y., & Gentry, M. (2018). A meta-analysis of gifted and talented identification practices. *Gifted Child Quarterly, 62*(2), 147–174. https://doi.org/10.1177/0016986217752107

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*(3), 299–324. https://doi.org/10.1111/j.1471-6402.1990.tb00022.x

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Kassambara, R. (2021). *rstatix: Pipe-friendly framework for basic statistical tests*. R package version 0.7.0 https://cran.r-project.org/web/packages/rstatix/index.html

Kovacs, K. & Conway, A. R. A. (2019). A unified cognitive/differential approach to human intelligence : Implications for IQ testing. (Journal of Applied Research in Memory and Cognition 8 255–272.

Lee, H., Karakis, N., Olcay Akce, B., Azzam Tuzgen, A., Karami, S., Gentry, M., & Maeda, Y. (2021). A meta-analytic evaluation of Naglieri Nonverbal Ability Test: Exploring its validity evidence and effectiveness in equitably identifying gifted students. *Gifted Child Quarterly, 65(3)*, 199-219.

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research, 41*(1), 63-76. https://doi.org/10.1080/0013188990410106

List, A., & Dykeman, C. (2019). The relationship of gifted program enrollment to race/ethnicity, gender, and SES. https://doi.org/10.31234/osf.io/zx396

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*, 111–138. https://doi.org/10.1177/001698620504900203

Lohman, D. F. (2011). *Cognitive Abilities Test (Form 7).* Riverside Publishing.

Lohman, D. F. (2012). *Cognitive Abilities Test (Form 7) Research and Development Guide.* Riverside Insights.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Riverside Publishing.

Lohman, D. F., & Lakin, J. M. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology, 79*(2), 389–407. https://doi.org/10.1348/000709908X354609

Luria, A. R. (1982). *Cognitive development. Its Cultural and Social Foundations*. Harvard University Press.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*(5), 481–498. https://doi.org/10.1016/j.intell.2004.06.008

McCallum, R. S., Bracken, B. A., & Wasserman, J. D. (2001). *Essentials of psychological assessment series. Essentials of nonverbal assessment*. John Wiley & Sons Inc.

McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology, 28,* 59–88. http://dx.doi.org/10.1080/15377903.2012.643757

McFadden v. Board of Education for Illinois School District U-46, No. 05 C0760 (2013).

Miller-Jones, D. (1989). Culture and testing. *American Psychologist, 44*(2), 360–366. https://doi.org/10.1037/0003-066X.44.2.360

Mills, C. J., Ablard, K. E., & Stumpf, H. (1993). Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills. *Journal of Educational Psychology, 85*(2), 340–346. https://doi.org/10.1037/0022-0663.85.2.340

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test.* The Psychological Corporation.

Naglieri, J. A. (2008). *Naglieri Nonverbal Ability Test–Second Edition (NNAT2)*. Pearson.

Naglieri, J. A. (2011). *Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT).* (D. Y. Ford, Ed.). In T. C. Grantham, D. Y. Ford, M. S. Henfield, M. T. Scott, D. A. Harmon, S. Porchèr, & C. Price (Eds.), *Gifted and advanced Black students in school: An anthology of critical works* (p. 243–252). Prufrock Press Inc.

Naglieri, J. A., Brulles, D., & Lansdowne, K. (2021). *Naglieri General Ability Tests*. Multi-Health Systems Inc.

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly, 49*, 29–36. https://doi.org/10.1177/001698620504900104

Naglieri, J. A. & Otero, T. M. (2017). *Essentials of CAS2 Assessment*. Wiley.

Naglieri, J. A. & Otero, T. M. (2018). The Cognitive Assessment System-2: From theory to practice. In Flanagan, D. P., & Harrison, P. L. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed.). Guilford Press.

Naglieri, J. A., Rojahn, J. & Matto, H. C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence*, *35*(6), 568–579. https://doi.org/10.1016/j.intell.2006.11.001

Naglieri, J. A., & Ronning, M. E. (2000). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 12,* 328-334.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of* Sciences, *106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Olszewski-Kubilius, P., & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly, 62*(1), 37–55. https://doi.org/10.1177/0016986217738015

Otis, A. S., & Lennon, R. T. (2003). *Otis-Lennon School Ability Test (OLSAT) 8th Edition: Technical Manual.* Harcourt Brace Educational Measurement.

Peters, S, J. & Engerrand, K. G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly*, 1-13.

Pintner, R. (1923). *Intelligence testing: Methods and results*. H. Holt.

Pfeiffer, S. I. (2012). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessment, 30,* 3–9. https://doi.org/10.1177/0734282911428192

Portes, A., & MacLeod, D. (1996). The educational progress of children of immigrants: The roles of class, ethnicity, and school context. *Sociology of Education*, *69*(4), 255-275. https://doi.org/10.2307/2112714

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ramsden, S., Richardson, F. M., Josse, G., Shakeshaft, C., Seghier, M. L., & Price, C. (2013). The influence of reading ability on subsequent changes in verbal IQ in the teenage years. *Developmental Cognitive Neuroscience, 6*, 30-9. https://doi.org/10.1016/j.dcn.2013.06.001

Randall, C. (2021). *The underrepresentation of twice-exceptional students in gifted programs.* [Unpublished doctoral dissertation]. Piedmont University.

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment.* Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4_11

Rinn, A. N., Mun, R. U., & Hodges, J. (2020). *2018-2019 State of the States in Gifted Education*. National Association for Gifted Children and the Council of State Directors of

Programs for the Gifted. https://www.nagc.org/2018-2019-state-states-gifted-education

Robinson, N. M. (2005). In defense of a psychometric approach to the definition of academic giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of Giftedness* (2nd ed., pp. 280–294). Cambridge University Press.

Rojahn, J., & Naglieri, J. A. (2006). Developmental gender differences on the Naglieri Nonverbal Ability Test in a nationally normed sample of 5-17 year olds. *Intelligence*, *34*, 253-260. https://doi.org/10.1016/j.intell.2005.09.004

Rstudio Team (2019). *Rstudio: Integrated Development for R*. Rstudio, Inc.

Ryan, J. J., & French, J. R. (1976). Long-term grade predictions for intelligence and achievement tests in schools of differing socioeconomic levels. *Educational & Psychological Measurement*, *36*, 553-559. https://doi.org/10.1177/001316447603600245

Sattler, J. M. (1988). *Assessment of children (3rd. ed.).* J. M. Sattler.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement.* Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson Tests of Cognitive Abilities, Fourth Edition*. Riverside Publishing.

Skiba, R. J., Knesting, K., & Bush, L. D. (2002). Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies*, *11*, 61-78. https://doi.org/10.1023/A:1014767511894

Steele J. (2003). Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology*, *33*, 2587- 2606. https://doi.org/10.1111/j.1559-1816.2003.tb02782.x

Strand, S., Deary, I .J., & Smith, P. (2006). Sex differences in cognitive ability test scores: A UK national picture. *British Journal of Educational Psychology*, *76*, 463-480. https://doi.org/10.1348/000709905X50906

Tyler-Wood, T., & Carri, L. (1993). Verbal measures of cognitive ability: The gifted low SES student's albatross. *Roeper Review*, *16*, 102-106. https://doi.org/10.1080/02783199309553550

US Census Bureau. (2016). *American Community Survey 2016: 1-year estimates*. Government Printing Office. Retrieved from https://www.census.gov

US Department of Education (2012). Civil Rights Data Collection. Retrieved from ocrdata.ed.gov

Wasserman, J. D., & Becker, K. A. (2000). Racial and ethnic group mean score differences on intelligence tests. Paper presented at the annual meeting of the *American Psychological Association*.

Warne, R.T., Anderson, B., & Johnson, A. O. (2013). The impact of race and ethnicity on the identification process for giftedness in Utah. *Journal for the Education of the Gifted, 36*, 487–508. https://doi.org/10.1177/0162353213506065

Worrell, F. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly*, *53*, 242–244. https://doi.org/10.1177/0016986209346828

Wright, B. L., Ford, D. Y., & Moore, J. L. (2022). *Creating Equitable Services for the Gifted*. IGI Global. DOI: 10.4018/978-1-7998-8153-7.ch002

Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. Henry Holt and Company.

Yoon, S. Y., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*, *53*, 121–136. https://doi.org/10.1177/0016986208330564