Personality and Individual Differences 104 (2017) 272–278

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Sex differences in achievement: Distributions matter

Jennifer R. Pargulski *, Matthew R. Reynolds

Educational Psychology, University of Kansas, Lawrence, KS, United States

ARTICLE INFO

Article history:
Received 16 May 2016
Received in revised form 6 August 2016
Accepted 12 August 2016
Available online xxxx

Keywords:
Sex differences
Academic achievement
Variance ratios

ABSTRACT

Sex differences in general and specific aspects of reading, math, and writing were investigated using the standardization sample from the Wechsler Individual Achievement Test–Third Edition (Pearson, 2009a). We examined mean and variance differences, differences at points along score distributions, and the percentage of males and females in the normative descriptive ranges used for the measures (e.g., average, high average). Statistically significant and small differences were found for Oral Reading Fluency, Math Problem Solving, and a Mathematics composite. Male advantages in Math Problem Solving were more substantial among high math achievers, and the male advantage on Mathematics was largely due to advantages on Math Problem Solving. Small to moderate sex differences were found for writing. Females scored consistently higher across those distributions, with the largest differences being in essay writing and sentence composition. Percentages of the sexes whose scores were in normative descriptive categories varied across measures. Although it has been concluded generally that male and female achievement is similar due to small effect sizes, those conclusions are limited as differences are often a function of achievement level. These differences have important real-world implications, for example, related to identifying those with learning problems and talents.

1. Introduction

An understanding of how and why sex differences in academic achievement arise requires an accurate description of those differences (or similarities). Findings are nuanced. Simple mean differences often hide important information that is revealed only after entire distributions of achievement scores are studied (Martin & Hoover, 1987; McGraw, Lubienski, & Strutchens, 2006; Penner & Paret, 2008). Compared to studies of mean differences; however, fewer studies have investigated sex differences across test score distributions–and in particular, in relation to writing. Further absent are large-scale studies of score distributions using nationally representative samples from well-developed, individually-administered achievement measures. These measures (e.g., Woodcock-Johnson, Wechsler and Kaufman tests) result from years of rigorous research, development, and standardization processes. They differ from group-administered tests because they are administered by a trained examiner to a single examinee to ensure high quality responses, thus minimizing sources of error, such as inattention and failure to understand instructions. Further, these instruments are designed to measure achievement across age and grade spans, allowing for measurement outside the bounds of a specific age or grade level.

Here we studied sex differences in achievement using standardization sample data from the Wechsler Individual Achievement Test-Third Edition (WIAT-III; Pearson, 2009a). Although we analyzed mean and variance differences in reading, math, and writing, we also tested for mean differences along the score distributions using quantile regression. Further, we calculated percentages of males in normative descriptive levels, which allowed for us to describe learning problems or talents may be identified across sexes.

1.1. Sex differences in reading, math, writing

Studies have shown slight advantages for males on mathematics tests and slight advantages for females on reading tests (e.g., Martin, Mullis, Gonzalez, & Kennedy, 2003; Nowell & Hedges, 1998). Effect sizes, however, are often trivial-to-small (see Hyde, 2014), and may depend on the specific skill measured (Martin & Hoover, 1987). For example, Hyde, Lindberg, Linn, Ellis, and Williams (2008) found no sex difference in lower-order math skills, and a small, but non-meaningful male advantage in higher-level math skills. Alternatively, moderate-to-large female advantages in writing have been found on United States national attainment tests of writing achievement (National Assessment of Educational Progress [NAEP], National Center for Education Statistics, 2012) and on individually-administered

* Corresponding author at: Educational Psychology, Joseph R. Pearson Hall, 1122 W. Campus Rd., University of Kansas, Lawrence, KS, 66045, United States.
E-mail addresses: jenpargulski@gmail.com (J.R. Pargulski), mreynolds@ku.edu (M.R. Reynolds).

http://dx.doi.org/10.1016/j.paid.2016.08.016

achievement tests (Camarata & Woodcock, 2006; Reynolds, Scheiber, Hajovsky, Schwartz, & Kaufman, 2015; Scheiber, Reynolds, Hajovsky, & Kaufman, 2015).

1.2. Achievement score distributions and sex differences

Means inform sex differences in achievement, but so does score variability. For example, male-to-female proportion ratios may differ in the tails of distributions due to differences in means, variances, or both (Nowell & Hedges, 1998). And although mean differences in reading and math have not been substantial, important sex differences in the tails have been documented (McGraw et al., 2006; Robinson & Lubienski, 2011). For instance, males have shown larger variability in mathematics in addition to slightly higher means (Feingold, 1992). As such, males are largely overrepresented among high math scorers.

Simple mean differences may also mask important systematic patterns within data. An informative analysis of sex differences includes differences along the score distribution. For example, McGraw et al. (2006) found that male advantages in NAEP math scores were largest at the upper tail of the distribution. Martin and Hoover (1987), using the group-administered Iowa Tests of Basic Skills, similarly found male advantages in math problem-solving that were most pronounced above the median of scores. They also found female advantages in Spelling and Reading Comprehension, but those were most pronounced below the median. Additionally, Penner and Paret (2008) found male advantages in the upper tail of math scores, and developmentally, those differences appeared sooner than did simple mean differences.

Differences at the "extremes" have important implications related to access to special services and programs of study. The few studies that have investigated sex differences using large-scale nationally-representative samples with individually-administered achievement tests have been limited mostly to means and, sometimes, variances (e.g., Scheiber et al., 2015). More fine-grained analyses with large-scale samples using these measures would contribute to the literature. These instruments assess wide ranges of achievement levels, allowing for differentiation at the tails, and they allow for better control of construct irrelevant variance because of individual administration. Further, these instruments assess the same student across achievement domains, including reading, math, and writing, allowing for comparisons across achievement areas.

1.3. Purpose

Previous research has examined mean and variance differences across sex in academic achievement; outside of math and sometimes reading, however, studies often forego analysis related to score distributions. To our knowledge, no recent large-scale studies have used standardization data from individually-administered achievement tests to examine sex differences along score distributions. Our purpose was to study sex differences in reading, math, and writing, while considering the score distributions, as well as specific and general aspects of reading, math, and writing, using data from the WIAT-III standardization sample.

We intended to answer the following questions:

1. Are there mean and variance differences across sex in reading, math, and writing scores?
2. Are there sex differences in reading, math, and writing at different points (i.e., different percentiles) along the distributions?
3. What percentage of males and females are in each of the normative descriptive ranges for reading, math, and writing (i.e., Extremely to Very Low, Low, Low Average, Average, High Average, High, and Very to Extremely High)?

2. Method

2.1. Participants

There were 2772 examinees, ages 4–19 years (preschool–12th grade), in the WIAT-III grade norm sample used in our study.[1] Standardization took place over Spring ($n = 1400$) and Fall ($n = 1375$) of 2008, with both Fall and Spring norms developed. This sample reflected demographic characteristics outlined in the 2005 United States Census, and was stratified on age/grade, sex, race/ethnicity, parent education level, geographic region, and special groups (e.g., specific learning disabilities) (Pearson, 2009b). Males and females were equally represented in our sample.

2.2. Measure

The WIAT-III is an individually-administered, diagnostic test of educational achievement (Pearson, 2009a). WIAT-III reading, math, and writing subtests and composites were used in this research.

2.2.1. Reading

The Total Reading composite and the Word Reading, Pseudoword Decoding, Oral Reading Fluency, and Reading Comprehension subtests were used.

Word Reading requires examinees to read a wordlist, without context. It measures word recognition skills. Internal reliability coefficients averaged across grades for Word Reading scores were 0.97 in both Fall and Spring. The test-retest reliability coefficient was 0.94.

Pseudoword Decoding requires examinees to read aloud from a list of pseudowords. The average internal reliability coefficient for Pseudoword Decoding scores was 0.96 in both Fall and Spring. The test-retest reliability coefficient was 0.94.

Oral Reading Fluency requires examinees to read passages aloud, measuring oral reading speed and accuracy. Only test-retest reliability (i.e., 0.94) was reported because there are no item level data.

Reading Comprehension requires examinees to read passages, and then respond to literal and inferential comprehension questions read aloud by the examiner. It measures Reading Comprehension of various text types (e.g., fictional stories, informational text, and how-to passages). Average internal reliability coefficients estimates were 0.88 and 0.86 in the Fall and Spring, respectively. The test-retest reliability coefficient was 0.90.

Last, Word Reading, Pseudoword Decoding, Oral Reading Fluency, and Reading Comprehension scores comprise the Total Reading composite. Average internal reliability coefficients estimated for Total Reading scores were 0.98 and 0.97 in the Fall and Spring, respectively. The test-retest reliability coefficient was 0.96 (Pearson, 2009b).

2.2.2. Math

The Mathematics composite and Math Problem Solving and Numerical Operations subtests were used in this research.

Math Problem Solving requires examinees to respond to questions presented orally and often with visual cues, and requires the application of math reasoning skills involving basic concepts, everyday applications, geometry, and algebra. Average internal reliability coefficients were 0.92 and 0.91 in the Fall and Spring, respectively. The test-retest reliability coefficient was 0.85.

Numerical Operations requires examinees to complete math calculation problems presented on a worksheet. It measures untimed written calculation skills involving basic skills, basic operations with integers, geometry, algebra, and calculus. Average internal reliability coefficients were 0.93 and 0.92 in the Fall and Spring, respectively. The test-retest reliability coefficient was 0.89.

[1] Three participants were deleted, two of whom were 3 years-old (one male and female) and one of whom was 20 years-old (male) to keep the age ranges as 4–19.