Q-interactive®

Equivalence of Q-interactive® and Paper Administration of Cognitive Tasks: WAIS®-5

Q-interactive Technical Report 16

Susan Engi Raiford, PhD

Troy Courville, PhD

September 2024

Pearson

Introduction

Q-interactive® is a platform that provides digitally assisted administration and scoring for individual clinical tests. Prior Q-interactive technical reports and equivalence studies provide extensive background about Pearson's digital adaptation approach and rationale. Cumulative results indicate that examinees respond in a similar way when stimuli are presented in Q-interactive rather than in a paper stimulus book, or when their tactile responses are detected and captured by the platform rather than in a handwritten format. Accrued evidence additionally indicates that when examiners use Q-interactive in place of record forms and stopwatches, they obtain the same results. All fifteen Wechsler Adult Intelligence Scale® (4th ed.; WAIS®–IV; Wechsler, 2008) subtests yielded comparable scores in Q-interactive (digital) and paper administration formats (Daniel, 2012a).

The present study evaluated the equivalence of scores from digital and paper formats of the Wechsler Adult Intelligence Scale (5th ed.; WAIS-5; Wechsler, 2024). Both subtest and composite score equivalence were appraised because they provide different evidence (i.e., task-level versus interpretive level).

WAIS-5 Equivalence Study

Study Design

Study designs that involve each examinee taking the test only once avoid readministration effects that can impact equivalence results because such designs approximate most closely a realistic testing experience. An equivalent-groups design, with random assignment of examinees to one of two conditions (i.e., digital format or paper format), was used for this study. Study designs commonly used to study equivalence are described in detail in Q-interactive Technical Reports 1 and 2 (Daniel, 2012a, 2012b).

An effect size of 0.20 or smaller is used as the standard for equivalence. The effect size is the average amount of difference between digital and paper format scores, divided by the standard deviation of scores in the population. An effect size of 0.20 is slightly more than one-half of a scaled-score point for subtests (which have means [*M*s] of 10 and standard deviations [*SD*s] of 3) and is three standard-score points for composites (which have *M*s of 100 and *SD*s of 15).

Method

This study was carried out as part of the WAIS-5 national tryout and standardization using the procedures for examinee recruitment, examiner selection and training, data collection, and scoring detailed in Chapter 3 of the *WAIS-5 Technical and Interpretive Manual* (Wechsler et al., 2024).

Participants

A sample of nonclinical examinees was recruited that was representative of the U.S. English-speaking population of individuals ages 16–90 with respect to education level, race/ethnicity, and sex according to 2022 U.S. Census Bureau data (Ruggles et al., 2023).

For Stage 1 of the study, we matched pairs of participants based on age, education level, race/ethnicity, and sex. One person from each pair was randomly selected for placement in the digital group. The other person was placed in the paper group.

The demographic characteristics of the digital group appear in Table 1.

Table 1. Demographics of the Digital Group

N	498
Age	
M	49.4
SD	24.6
Range	16–90
Education	
Less than or equal to 8 years of school	4.8
9–12 years of school, no diploma	7.6
High school diploma or equivalent	24.9
Some college or technical school, associate degree	29.7
Bachelor's degree	32.9
Race/Ethnicity	
African American	13.7
Asian	4.8
Hispanic	17.7
Other	4.6
White	59.2
Sex	
Female	68.9
Male	31.1

Note. Except for sample size (*N*) and age, data are reported as percentages. For an examinee ages 16–24, education is based on the average parents' education; for an examinee ages 25–90, education is based on the examinee's education.

For Stage 2 of the study, we randomly selected 500 groups from the WAIS-5 nonclinical paper samples. Each of these 500 groups (i.e., 500 matched paper groups) were chosen to match the education levels and sample size of the digital group described previously. All examinees included met the criteria to be included in the normative sample. Details about the normative sample criteria and selection can be found in Chapter 3 of the *WAIS-5 Technical and Interpretive Manual* (Wechsler et al., 2024).

Measure

The WAIS-5 is an individually administered, comprehensive clinical instrument for assessing the intelligence of adolescents and adults ages 16–90. Many of its subtests are nearly identical in administration and scoring to those of the WAIS–IV, which was already evaluated for equivalence of digital and paper formats (Daniel, 2012a). Two of the new subtests from the Wechsler Memory Scale® (4th ed.; WMS®–IV; Wechsler, 2009), Symbol Span and Spatial Addition, were also already evaluated for equivalence (Daniel, 2013). Because some subtests are new, and others have had minor changes in administration or scoring (including changes to the Q-interactive examiner interface), the entire battery has been re-evaluated for equivalence.

The digital group was administered the WAIS-5 standardization version in digital format on Q-interactive Assess Version 1. The paper group and the 500 matched paper groups were all administered the WAIS-5 standardization version in paper format.

For the WMS–IV version of Spatial Addition, the examinee used paper components (i.e., cardboard memory grid and cards) to respond. A new digital version of Spatial Addition for the WAIS-5 and Wechsler Memory Scale (5th ed.; WMS-5; Wechsler, in press) was administered in which the examinee responds by touching and dragging digital objects on the tablet. In order to minimize construct-irrelevant differences between the digital and paper formats, the paper materials were modified (i.e., chips and flat grid) for the WAIS-5 to resemble the digital format. Nevertheless, it was assumed that the new version of Spatial Addition would not be raw-score equivalent in digital and paper formats because of the difference in response mode. The goal was for the norms of the two Spatial Addition formats to be equated so they could be treated as alternate forms for purposes of clinical interpretation.

Procedure

Examiners captured response information in the standard manner used for norming, which includes writing the complete verbatim response to each verbal subtest item. The Pearson research team checked paper- and digital-administration cases for proper use of queries, and paper-administration cases for other administration rules (e.g., start points, reversal rules, discontinue rules). All subtest raw scores were calculated automatically, either by Pearson staff using the keyed item-level data or by the Q-interactive system.

The data analysis was conducted in two stages.

Stage 1

Stage 1 focused on examining the effect size of the mean score differences between the digital and paper groups. The predetermined effect size criterion of 0.20 or smaller was used to indicate equivalence. Any subtest or composite with an effect size greater than 0.20 would be considered not equivalent. In this situation, the norms of that subtest or composite would be adjusted using equating techniques or the subtests involved reexamined for explanations. It was expected a priori that Spatial Addition would not show equivalence and that the norms would require adjustment.

Stage 2

Stage 2 involved a deeper examination of the subtest and composite scores at all score levels of the distributions. To accomplish this, data gathered from the 500 matched paper groups from the nonclinical samples were used to establish a range of cumulative percentages of scores (i.e., confidence intervals) observed for each scaled and standard score level in the distribution.

If administration of the WAIS-5 in either format results in equivalent scores, the digital group's scores should generally fall within the confidence intervals that were established using the 500 matched paper groups. Cumulative frequency tables and cumulative distribution function plots of the subtest and composite score distributions were visually inspected to evaluate the nature of the digital score distributions in relation to the typically observed paper score ranges.

Results

Stage 1

Table 2 reports the *M*s and *SD*s of scores for the digital and paper groups and the standard differences (Cohen's *d*) between the groups. For subtests that previously appeared in equivalence studies for the WAIS–IV (Daniel, 2012a) and WMS–IV (Daniel, 2013), Cohen's *d* is provided in the far-right column.

Table 2. Mean Performance of Digital and Paper Groups

		Administration format						
Subtest/Composite Score	*n*	Digital *M*	Digital *SD*	Paper *M*	Paper *SD*	Difference	Cohen's *d*	WAIS–IV *d*
Verbal Comprehension	479	99.3	14.7	99.4	14.1	-.07	.00	
Visual Spatial	495	98.9	14.3	99.6	14.1	-.70	-.05	
Fluid Reasoning	481	99.8	14.2	99.9	14.9	-.09	-.01	
Working Memory	487	99.6	15.4	99.4	14.0	.14	.01	
Processing Speed	329	100.7	15.1	101.5	15.2	-.83	-.05	
Full Scale IQ	317	100.4	14.4	100.7	14.3	-.30	-.02	
Verbal (Expanded Crystallized)	479	98.7	15.1	98.9	13.9	-.15	-.01	
Verbal Reasoning	485	98.7	14.9	99.6	14.3	-.89	-.06	
Expanded Visual Spatial	490	98.5	14.9	99.5	14.0	-.99	-.07	
Expanded Fluid	481	99.0	14.0	99.9	14.3	-.93	-.07	
Quantitative Reasoning	484	97.9	13.9	99.4	14.2	-1.56	-.11	
Expanded Working Memory	481	99.4	15.1	99.5	13.8	-.05	.00	
Visual Working Memory	490	98.3	15.5	99.6	13.9	-1.25	-.08	
Auditory Working Mem-Registration	487	98.3	15.0	99.5	14.8	-1.21	-.08	
Auditory Working Mem-Manipulation	494	100.2	15.3	99.7	14.0	.44	.03	
Expanded Processing Speed	468	97.3	14.3	100.0	15.2	-2.70	-.18	
Motor-Reduced Processing Speed	480	98.9	14.5	99.5	14.8	-.60	-.04	
Nonverbal	317	99.7	14.4	100.8	14.2	-1.12	-.08	
Nonmotor	467	99.5	14.2	99.7	14.1	-.20	-.01	
General Ability	467	100.0	14.2	99.6	14.4	.40	.03	
Cognitive Proficiency	323	99.7	15.5	100.8	14.2	-1.10	-.10	
Similarities	485	9.9	3.1	9.9	3.0	.06	.02	-.11
Vocabulary	481	9.8	2.9	9.9	2.8	-.08	-.03	-.05
Information	487	9.8	2.9	9.6	2.9	.11	.04	-.12
Comprehension	487	9.6	3.1	9.8	2.8	-.24	-.08	-.12
Block Design	495	10.0	2.9	9.9	2.9	.13	.04	-.16
Visual Puzzles	495	9.6	2.9	10.0	2.8	-.38	-.13	.18
Matrix Reasoning	490	10.1	3.0	10.0	3.0	.13	.04	.10
Figure Weights	484	9.7	2.9	9.9	3.0	-.21	-.07	-.02
Arithmetic	495	9.4	2.9	9.8	2.9	-.37	-.13	.04
Set Relations	495	9.8	3.0	10.1	2.9	-.32	-.11	
Digit Sequencing	495	10.0	3.0	9.9	2.8	.11	.04	
Running Digits	487	9.8	3.1	9.9	3.0	-.06	-.02	
Digits Forward	495	9.6	3.0	9.9	3.0	-.37	-.12	
Digits Backward	494	9.9	2.9	10.0	2.9	-.12	-.04	
Symbol Span*	493	9.6	2.9	10.0	2.9	-.37	-.13	-.06
Spatial Addition*	492	8.9	3.2	10.0	2.9	-1.05	-.34	.11
Spatial Addition after equating	492	9.9	3.2	10.0	2.9	-.08	-.03	
Letter-Number Sequencing	495	10.2	3.2	10.0	2.9	.20	.07	-.04
Coding	331	10.3	3.0	10.4	3.1	-.05	-.02	.07
Symbol Search	337	10.0	3.1	10.1	3.0	-.18	-.06	.13
Naming Speed Quantity	495	9.6	2.9	9.8	3.0	-.25	-.08	

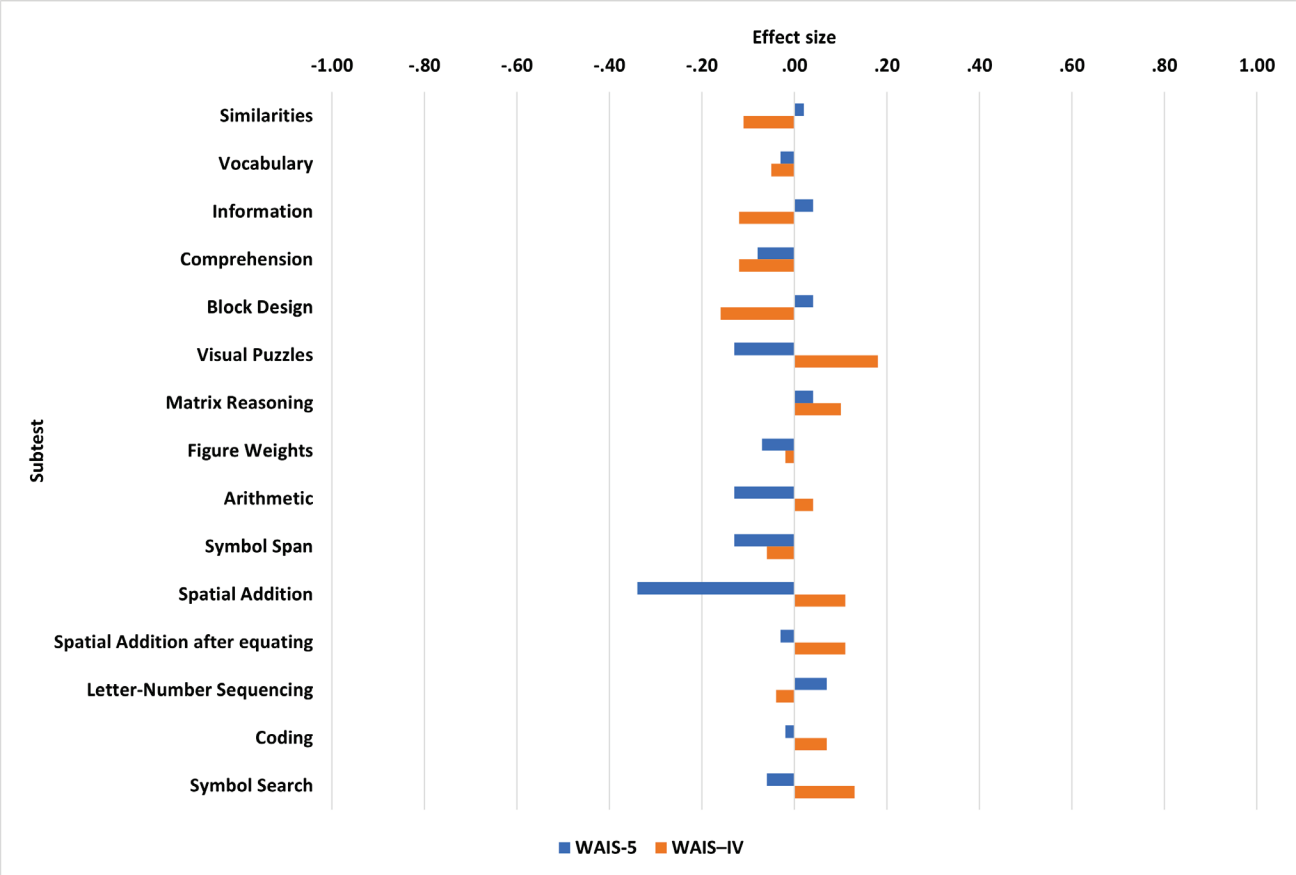
Note. A positive effect size indicates higher scores with digital format, and a negative value indicates lower scores with digital format.
The *n* counts for Coding and Symbol Search and composite scores that include them are lower due to missing data after changes to the digital format of these subtests for standardization.
* Symbol Span and Spatial Addition were WMS–IV subtests. Results shown for those subtests originate from Daniel (2013).
The mean Expanded Visual Spatial and Visual Working Memory Index scores are derived using Spatial Addition after equating.

As shown, the means of the digital and paper groups are highly similar, and the effect sizes of all composite score group mean differences are negligible. Similarly, the effect sizes for all subtest scaled score mean differences, with the exception of Spatial Addition, are negligible. Therefore, with the exception of Spatial Addition, the same WAIS-5 norms can be applied whether the subtests are administered using digital or paper format.

The Spatial Addition mean difference shows a small effect size prior to norms adjustment. After the norms for the digital format of Spatial Addition are adjusted using equating techniques, the mean difference effect size is negligible.

Figure 1 summarizes the effect sizes of the mean score differences observed for subtests that are common to both the WAIS-5 (in the present study) and the WAIS–IV (in Daniel, 2012a) or the WMS–IV (in Daniel, 2013).

Figure 1. Summary of WAIS-5 and Prior Equivalence Study Results



Note. Symbol Span and Spatial Addition were WMS–IV subtests. Results shown for those subtests originate from Daniel (2013).

As shown, the pattern of these negligible effect sizes is not similar across the WAIS-5 study and prior studies (r = -.43). The mean score difference effect sizes from the WAIS-5 study are frequently in the opposite direction and/or not of similar magnitude as those from the prior studies. These cumulative results suggest that format does not play a role in mean score differences observed between groups administered the WAIS-5 in differing formats.

Stage 2

Examples of the smoothed subtest cumulative distribution function plots for the digital sample and the confidence interval of random paper samples appear in Figure 2. One example of a subtest from each cognitive domain is provided for the sake of illustration. Spatial Addition (before and after equating) are also pictured to illustrate the benefit of adjusting the norms where needed. For each plot, the range of scaled scores (i.e., 1–19) is represented along the X-axis, with the mean scaled score (i.e., 10) occurring halfway across the axis. The cumulative percentage of scores at a given score or lower (i.e., 0–100) is represented along the Y-axis. The digital group's values are shown in red, and the paper group's values are shown in solid blue. The minimum and maximum values for the 500 matched paper groups are shown in dashed blue.

Figure 3 illustrates similar plots for the five primary index scores and the FSIQ, with ranges of 45–155 for the index scores and 40–160 for the FSIQ represented along the X-axis.

Figure 2. Examples of WAIS-5 Subtest Smoothed Cumulative Percentage Plots

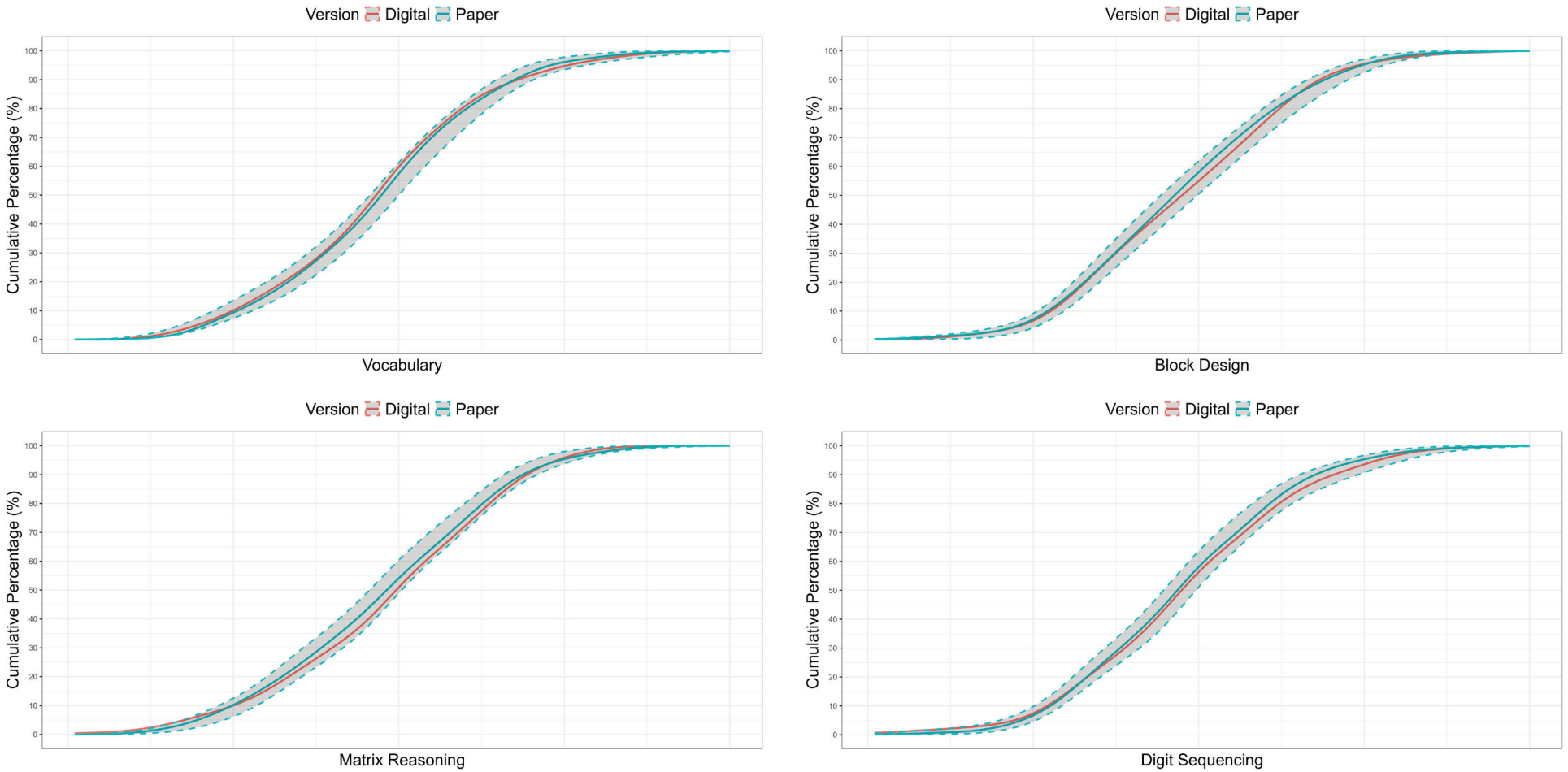
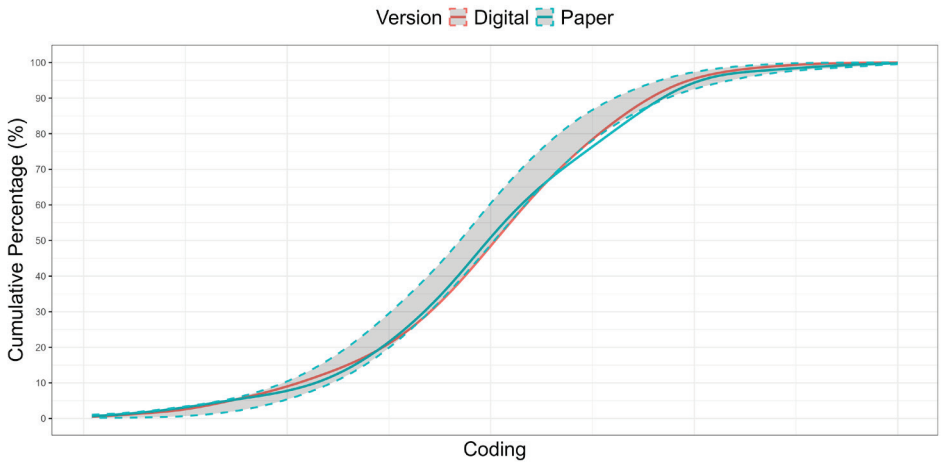
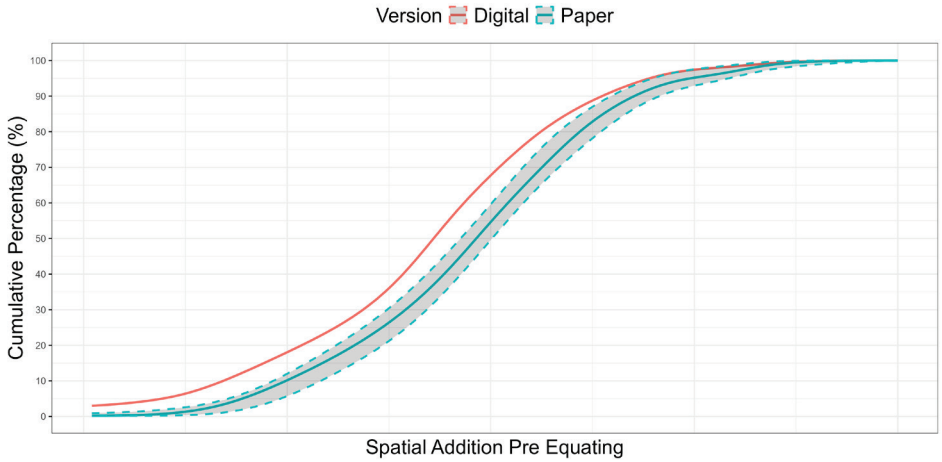


Figure 2. Examples of WAIS-5 Subtest Smoothed Cumulative Percentage Plots (*continued*)





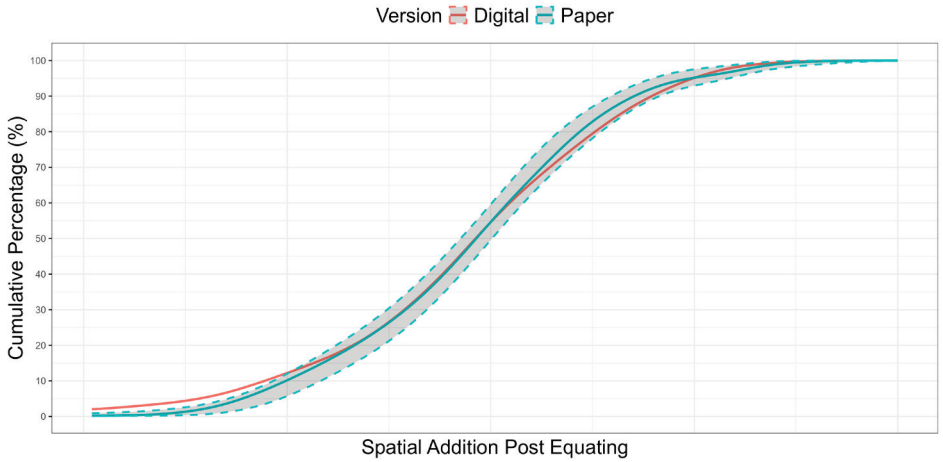


Figure 3. WAIS-5 Primary Index Scores and FSIQ Smoothed Cumulative Percentage Plots

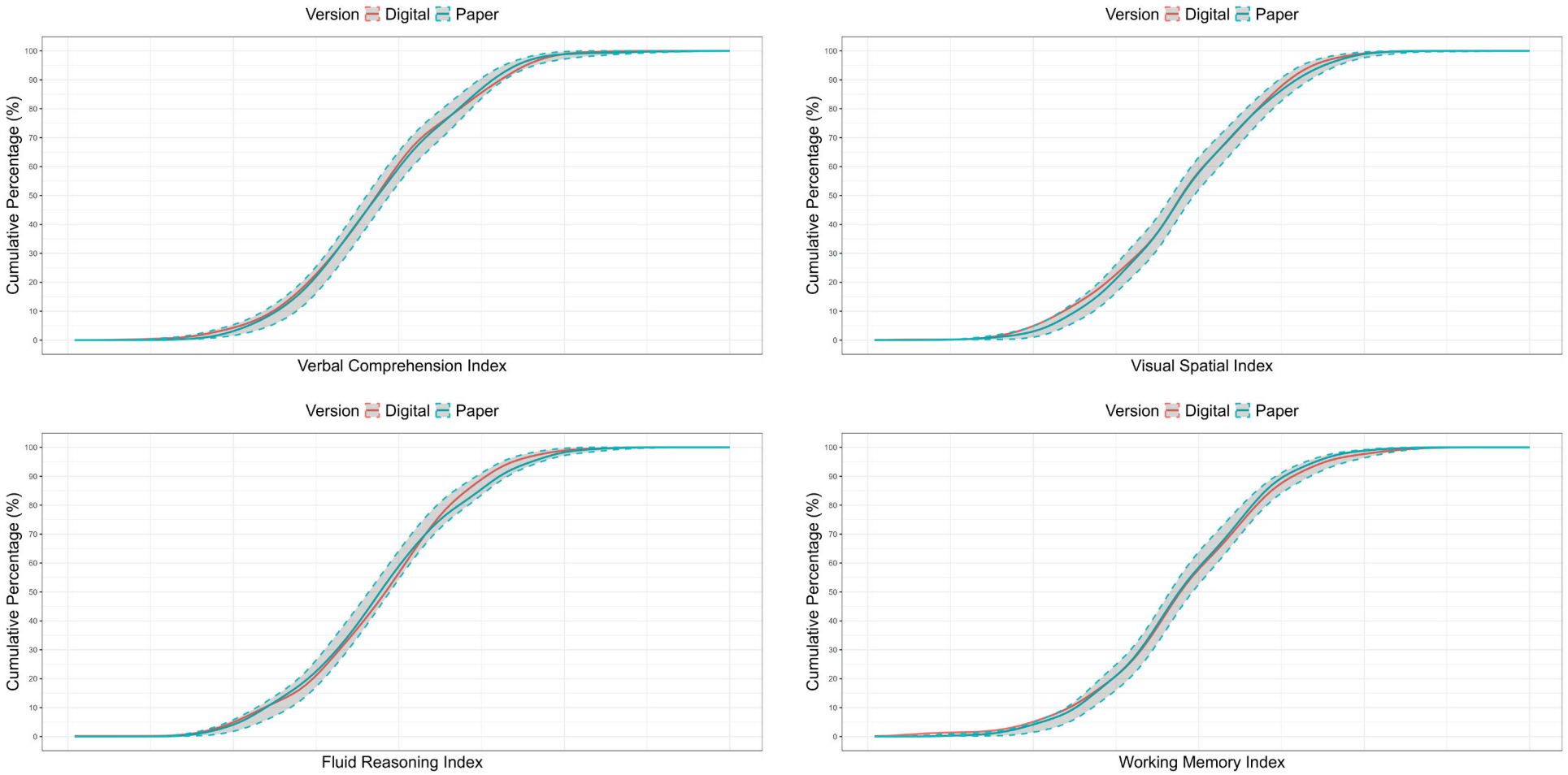
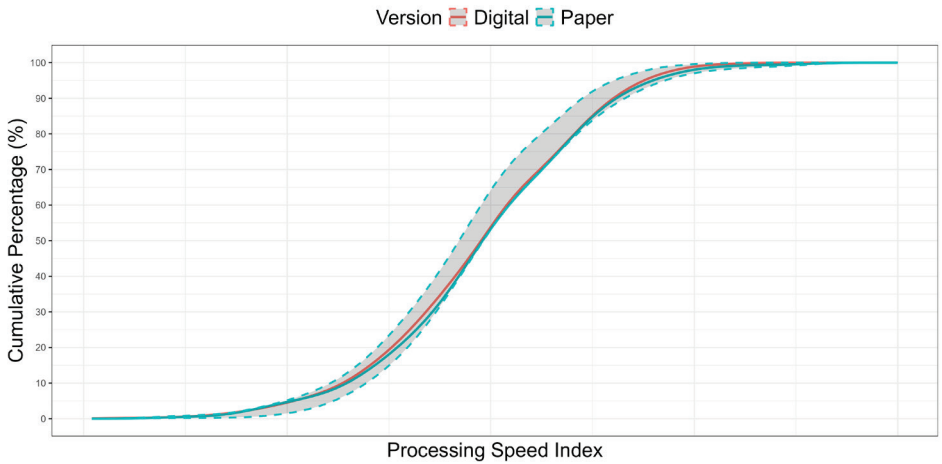
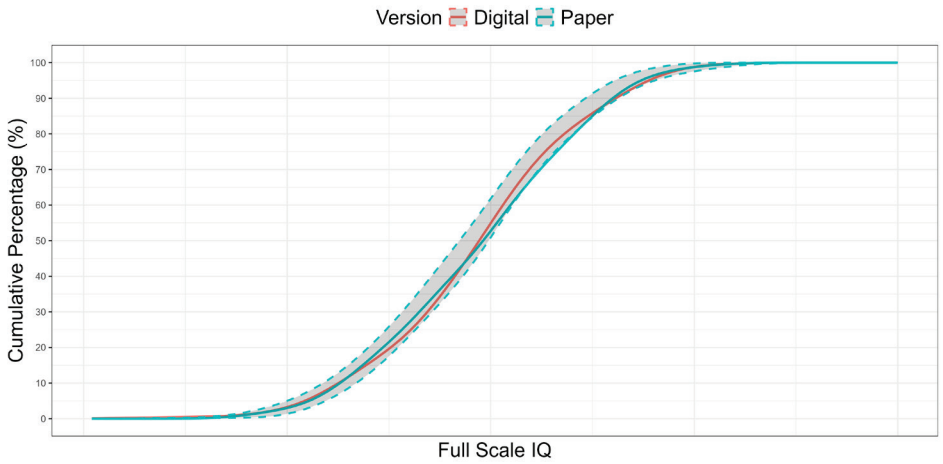


Figure 3. WAIS-5 Primary Index Scores and FSIQ Smoothed Cumulative Percentage Plots (*continued*)





As shown in Figures 2 and 3, the digital group's distributions generally fall within the confidence intervals. Inspection of the plots for the other subtests and for the ancillary composite scores indicate that all other digital group distributions are highly similar to those of the 500 matched paper groups at all levels.

Discussion

This study was an extension of prior equivalence studies (Daniel, 2012a, 2013) for 14 of the WAIS-5 subtests, as well as the initial equivalence study of six additional subtests. Format effect sizes for all of the subtests (with the exception of Spatial Addition before it was equated) fell within the established criterion for equivalence used in studies of Q-interactive (i.e., 0.20 or less). The cumulative distribution functions indicate the digital group's performance falls within the expected range of demographically matched paper groups and that the composite and subtest norms do not require adjustment at any level, with the exception of Spatial Addition.

The Spatial Addition digital format norms were adjusted using equating techniques. The adjusted norms for Spatial Addition in digital format are automatically applied in Q-interactive. The same is true when Q-interactive calculates the expanded Visual Spatial Index, the Expanded Working Memory Index, and the Visual Working Memory Index. Q-interactive users do not manually apply any conversion to the Spatial Addition norms for this reason.

References

Daniel, M. H. (2012a). *Equivalence of Q-interactive administered cognitive tasks: WAIS–IV. Q-interactive Technical Report 1.* Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/007-s-QinteractiveTechnical%20Report%201_WAIS-IV.pdf

Daniel, M. H. (2012b). *Equivalence of Q-interactive administered cognitive tasks: WISC–IV. Q-interactive Technical Report 2.* Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/009-s-Technical%20Report%202_WISC-IV_Final.pdf

Daniel, M. H. (2013). *Equivalence of Q-interactive and paper administration of WMS–IV cognitive tasks. Q-interactive Technical Report 6.* Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/0010-s-Technical_Report_6_WMS-IV.pdf

Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., Richards, S., Rogers, R., & Schouweiler, M. (2023). IPUMS USA (Version 14.0) [Data set]. IPUMS. https://doi.org/10.18128/D010.V14.0

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). NCS Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale* (4th ed.). NCS Pearson.

Wechsler, D. (2024). *Wechsler Adult Intelligence Scale* (5th ed.). NCS Pearson.

Wechsler, D. (in press). *Wechsler Memory Scale* (5th ed.). NCS Pearson.

Wechsler, D., Raiford, S. E., & Presnell, K. (2024). *Wechsler Adult Intelligence Scale* (5th ed.): *Technical and interpretive manual*. NCS Pearson.