



Q-interactive: Training Implications for Accuracy and Technology Integration

Stephanie Corcoran¹

Accepted: 15 February 2021 / Published online: 1 March 2021

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Abstract

With the iPad-mediated cognitive assessment gaining popularity with school districts and the need for alternative modes for training and instruction during this COVID-19 pandemic, school psychology training programs will need to adapt to effectively train their students to be competent in administering, scoring, and interpreting cognitive assessment instruments. This manuscript describes a mixed methods study of graduate students learning both the traditional and digital format (Q-interactive) of the WISC-V, with the goal of improving training methods and reducing administration and scoring errors. Results indicated that more errors are made on the traditional format than on the digital format, but the errors that did occur on the digital format were on subtests that require clinical acumen. Q-interactive did not reduce errors related to more complex judgments and nuanced scoring. The participating graduate students were surveyed regarding their perceptions of each format, and they revealed a majority preference for the digital format. Training implications are discussed, and specific suggestions provided for how training programs may respond to our current situation by integrating Q-interactive into their assessment courses.

Keywords Cognitive assessment · School psychology · Training · Digital assessment · Q-interactive · COVID-19

Practicing school psychologists devote approximately half of their time to assessment and two-thirds of their time to special education eligibility determination (Castillo, Curtis, & Gelley, 2012; Curtis, Hunley, & Grier, 2002; Fagan & Wise, 2007; Hosp & Reschly, 2002), including administration, scoring, interpretation, and reporting of the results of intelligence tests. Accordingly, school psychology training programs should be preparing trainees who demonstrate a knowledge of and skills in psychological and educational assessment (NASP, 2010b). The large majority of school psychology graduate training programs have at least one required course devoted to teaching these skills (Ready & Veague, 2014; Sotelo-Dynega & Dixon, 2014). These assessment courses develop essential diagnostic skills, familiarize students with the assessment process, and foster a sense of professional identity (Oakland & Jimerson, 2006). The assessment courses have demonstrated having a long-lasting impact, with professionals' practice remaining generally consistent with their training (Alfonso,

LaRocca, Oakland, Spanakos, 2000; Sotelo-Dynega & Dixon, 2014; Wilson & Reschly, 1996).

Traditionally, the Wechsler scales have been the preferred test for use with all of the disability categories (Styck & Walsh, 2016). Lockwood and Farmer (2019) found in their recent survey that 95% of instructors require at least one administration of the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) in assessment courses. Therefore, most programs provide training in the WISC-V, despite it being more difficult to administer and score compared to other tests such as the Woodcock-Johnson (Ramos & Alfonso, 2009). A common concern for trainers is the high prevalence of administration and scoring error rates, and with good reason, based on the evidence.

Errors

Results from research examining the occurrence of examiner errors on cognitive tests “indicate that examiner errors impact Index and FSIQ scores at alarmingly high rates” and that “examiner errors were more likely to spuriously inflate FSIQ scores than artificially depress them” (Styck & Walsh, 2016, p.13). The most common errors include failure to administer sample items, incorrect calculation of raw scores, failure to

✉ Stephanie Corcoran
corcoran@uab.edu

¹ School of Education/Department of Curriculum and Instruction, University of Alabama at Birmingham, 1150 10th Ave South, Birmingham, AL 35294, USA

record responses verbatim, and failure to query. In particular, a recent examination of 295 Wechsler protocols completed by graduate students and practicing school psychologists revealed that errors tend to be the norm, not the exception (Oak, et al., 2019). This finding is disturbing given that the results of these cognitive assessments inform high-stakes decisions, such as special education classification, gifted education classification, educational placement, and occasionally even death penalty determination (McDermott, et al., 2014). Even a one-point scoring error can be the difference between someone receiving services or being denied services. This line is especially sharp when using any form of ability or achievement assessment for students being considered for specific learning disabilities (SLD), based on cut-off scores. The legitimacy of these high-stakes decisions depends entirely on the accuracy of test scores (McDermott et al., 2014).

Given that graduate school training lays the foundation for future success with test administration and scoring, these errors have attracted the attention of researchers. Specifically, they have investigated the errors made by graduate students learning to administer the Wechsler scale. In one study, errors were committed on 98% of all protocols, and, as a result, the Full-Scale IQ was inaccurate on two-thirds of all protocols (Styck & Walsh, 2016). Further, graduate students did not significantly improve over three practice administrations despite receiving individualized written feedback about errors after each administration. Another study found no significant decrease in WISC-V errors over six administrations (Mrazik, et al., 2012). Simons, Goddard, and Patton (2002) reviewed the literature on examiner errors and came to the conclusion that errors are almost a given, writing, “the percentage of errors that can be expected when psychometric tests are scored by hand will be significantly greater than zero” (p. 297). The proportion of test protocols that contains at least one examiner error has been reported to range from 14 to 100% (Alfonso, et al., 1998; Allard & Faust, 2000; Slate & Chick, 1989), and these types of errors have been documented to result in changes to the full scale score for anywhere between 11 and 88% of examinees once errors on protocols were corrected (Alfonso et al., 1998; Belk et al., 2002; Slate, Jones, & Murray, 1991). By any measure, these numbers are concerning.

Moreover, studies show that it is not just graduate students or novice practitioners who frequently make errors (Belk et al., 2002; Gurley, 2008), but recurrent errors have also been found in protocols of practicing psychologists (Charter, Walden, & Padilla, 2000). Research suggests high rates of administration and scoring errors persist into independent practice (Styck & Walsh, 2016). It has been demonstrated that experienced examiners actually become more practiced at making errors (Belk, et al., 2002; Loe, et al., 2007).

This research highlights the need for improved training methods at both the graduate school level and in professional

development. Several studies (i.e., Oak et al. 2019; Slate et al., 1993) recommended additional training components, including the following: (a) reviewing the most common sources of administration and scoring errors during instruction, (b) providing specific feedback about the type and number of errors committed, (c) having students review each other’s administrations and scoring, and (d) requiring more than one video administration of the WISC to be reviewed by the instructor. Research has suggested that more passive approaches (students learning from their previous mistakes) have not significantly reduced student errors. Although graduate students are believed to be highly motivated, this motivation alone does not translate into decreasing errors with repeated administrations. Even punitive approaches (penalizing errors with reduced grades or re-dos) did not yield significant reductions in administration and scoring errors (Oak et al., 2019). In a recent nationwide survey by Lockwood & Farmer (2019), trainers reported that their students are required to complete video-taped administrations (78%), unobserved administrations (55%), administrations with an observer physically present (44%), and audio-taped administrations (4%). In addition, 98% report completing protocol reviews, and 33% have students complete simulated administrations. The majority of instructors surveyed (86%) did not currently require any tablet-based administrations. However, this survey was conducted prior to the COVID-19 pandemic.

The COVID-19 pandemic has changed the way school psychologists’ practice. The federal government has not waived the requirements under the IDEA and mandated assessments need to be completed (US Department of Education, 2020). Some school districts have responded by providing in-person testing in one on one settings using safety precautions determined by their given districts (clean rooms, plexiglass dividers, masks, and shields). Some school districts, even though they are providing 100% remote learning, are allowing in person assessments through “appointments.” Other districts are attempting remote assessments. Farmer et al. (2020) suggest caution in moving forward, if at all, with remote testing due to the unique challenges presented during the COVID-19 pandemic including lack of legal clarity and questions as to whether it is even possible to obtain reliable and valid assessment results. Although the circumstances and models vary widely from district to district, many school districts are moving toward increased use of technology and tablet-based testing. The increased use of tablets helps in the current COVID-19 environment because it utilizes less shared test materials that must be cleaned and disinfected between test sessions. Additionally, social distancing can be more easily implemented with the tablet administrations as the tablets work through Bluetooth at a safe distance of six feet or more. Therefore, adherence to CDC guidelines can be more easily maintained.

Q-interactive

In 2012, Pearson introduced a new technology-based testing product, Q-Interactive, which is a 1:1 iPad-based testing system that helps administer, score, and report 20 different clinical assessments, including the WISC-V. Q-interactive requires a computer with access to the internet, two tablets connected via Bluetooth, and a Q-interactive account and software (Cayton et al., 2012). According to Cayton et al. (2012), Pearson adapted currently existing and frequently utilized measures such as the WISC-V and WIAT-III to a digital format maintaining their original design elements. The goal was to make the assessment process simpler and more efficient by reducing the amount of time required for testing and make administration and scoring easier, more accurate, and convenient. Everything is included in one location, a tablet, rather than multiple materials (e.g., protocol, administration manual, notepad, or stopwatch). Krach et al. (2019) reported that a set of tablets containing two complete tests may weigh 5 lbs, where the equivalent paper version with all of its materials might weigh 20 lbs or more. This is no small consideration for individuals who must daily transport test materials into and out of school buildings. Q-interactive has consistent design elements with the traditional format. Scheller (2013) noted that the digital format of the Q-interactive also reduces the amount of information the examiner needs to memorize or refer back to during testing, including start points, discontinuation rules, and timing tasks. The digital system allows for automated scoring that follows the necessary rules while retaining flexibility, with the examiner maintaining control over the testing session (Cayton et al., 2012). Some computerized assessments, such as those frequently used in neuropsychology, are computer directed, and the examinee interacts with a computer or device without supervision or test observation by an examiner. The Q-interactive system, however, is examiner-directed keeping control in the hands of the examiner, with human interaction being an integral part of the process (Scheller, 2013).

Q-interactive Concerns

Despite its attractiveness to trainers and widespread use by school districts, there are notable concerns about Q-interactive. Most notable are the lack of psychometric equivalency between instrument versions and the lack of non-biased research, as Krach (2019) discusses in her recent study. Krach (2019) asserts that the little research that does exist on Q-interactive's psychometric equivalency has been conducted in-house and may be biased. In more recent research, Gilbert, Kranzler, and Benson (2020) conducted an independent study that demonstrated Q-interactive, and the traditional format are not equivalent. In their study, the authors found that

Q-interactive produced higher FSIQ and PSI scores, largely due to differences in performance on the "Coding" subtest. They urged examiners to use the traditional format of the WISC-V rather than the Q-interactive. They suggested that if you must use Q-interactive, to substitute the "Symbol Search" subtest for the "Coding" subtest in the calculation of FSIQ, to administer the PSI subtests with the traditional paper response booklets, and interpret the General Ability Index, which does not include a PSI subtest, instead of the FSIQ. Pearson subsequently made the recommendation to use the paper-based response booklets rather than use the tablet for Coding and Symbol Search.

Training Issues for Integrating Technology in Cognitive Assessments

The integration of technology in cognitive assessment has significant ethical implications. Training programs have the responsibility to teach their students the central principles of measurement from the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) and emphasize adherence to legal and ethical standards of practice (NASP, 2010a, 2010b). However, training specifically in computerized testing is limited, and students should be exposed to ethical concerns, potential judgment errors, and possible pitfalls in evaluating computer-generated reports (Shulenberg, & Yutzenka, 2004). Key ethical domains for trainers to address relate specifically to competence, utilization of computerized test interpretation packages, equivalency between traditional and computerized assessment procedures, confidentiality, and cultural, experiential, and disability factors (Shulenberg, & Yutzenka, 2004).

Students should be taught to become critical consumers of computer-based assessment (Snyder, 2000). The field of neuropsychology has long utilized computer-directed assessment and provide insight in these issues. The joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology (Bauer, Iverson, Cemich, Binder, Ruff, & Naugle, 2012) provides guidance to promote accurate and appropriate use of computerized tests that maximizes clinical utility and minimizes risks of misuse. Although their assessments tend to be more computer directed rather than examiner driven, the same core principles apply.

Privacy Protection and Data Integrity

Butcher (2003) states that the danger of misusing data applies to all psychological formats, but the risk seems particularly high when one considers the convenience of computerized outputs. Maintaining confidentiality, privacy, and security is

the responsibility of the professional who collects, stores, and transmits data, and they should have appropriate knowledge about information technology that assures client rights are protected (Bauer et al., 2012). Upholding this responsibility requires knowledge and understanding about how the data is collected, protected, and stored.

The frequency of human error in administration and scoring is recognized, and computers are generally perceived as superior in this regard (Shulenberg, & Yutzenka, 2004). However, Weiner and Greene (2007) suggested that only 60% of computer interpretations were clinically appropriate. Butcher (2003) cited a number of sources suggesting that computer-based reporting services may produce erroneous results. Trainers must emphasize to students that computer-based reports should not be used to replace clinical judgment but, rather, as aides to clinical interpretations (Bauer et al., 2012). They must ensure that the computerized interpretation is an accurate reflection of the examinee and not just a series of interpretations put forth in a software package by a testing company (Butcher, 2003). Snyder (2000) notes that use of computerized interpretive reports is risky and that examiners may be liable if they include a computer-generated interpretation in a report without adequately evaluating its relevance. Noland (2017) cautioned that fostering early reliance on the software at the expense of a deeper understanding of standardized test administration in general is unwise.

Cultural, Experiential, and Disability Factors

In addition to prioritizing awareness of issues relating to privacy and data integrity, trainers need to make students aware of the array of potentially confounding variables that may differentially influence the process, outcome, reliability, and validity of a computerized testing session (Schulenberg & Yutzenka, 2004). Foxcroft and Davies (2006) discussed the need to consider equality of access for all groups and the impact that inequality of access to computers and technology can have on test performance. Attitudes and feelings toward computers and their familiarity with them can affect test performance. Also, ensuring that appropriate normative information is available allows the examiner to determine if the test can be given to examinees from different racial, ethnic, and disability backgrounds (Bauer et al., 2012). Some examinees with cognitive, language, motor, or sensory issues might have difficulty completing a computerized test in the manner intended by the test developers. These and other ethical issues should be considered when selecting the test format.

Due to the newness of the Q-interactive technology, trainers lack research that examines implications for school psychology trainers involved in teaching graduate students to administer, score, and interpret cognitive assessments. Trainers need more information about the Q-interactive technology, specifically, evidence for why it should, or should not

be, taught and specific strategies for training students in using this type of technology. Additionally, trainers need to know whether Q-interactive increases the accuracy of assessments administered. A mixed methods study of graduate students learning both the traditional and digital formats of the WISC-V was designed to address these practical and research gaps in our knowledge.

The overarching research questions in this study are the following:

- 1) Does Q-interactive alleviate some of the administration and scoring errors that are common to students learning these tests?
- 2) What are graduate student's perceptions of traditional vs digital assessment formats?

Purpose

Cognitive assessment plays an essential role in the functioning and professional identity of school psychologists (Benson et al., 2019; Sotelo-Dynega & Dixon, 2014), and training programs have the responsibility to prepare their students to use cognitive assessments effectively for the purpose of sound data-based decision-making (American Psychological Association, 2015; NASP, 2010a, b). The purpose of this article is to investigate the iPad-based testing system, called Q-interactive, with the goal of improving training methods and reducing administration and scoring errors. Analysis of the error rates of traditional format vs digital format will be discussed as well as graduate student's perspectives on the traditional vs digital assessment formats. Answers to these research questions and current relevant issues will assist trainers in determining whether to provide training in Q-interactive, and how to do it most efficiently and effectively.

Methods

Study Design

With Institutional Research Board (IRB) approval, data were collected across three semesters from graduate students enrolled in the cognitive assessment course in a School of Education at a research-intensive university in the southern portion of the USA. Participation in completing the follow-up survey was optional, and the electronic surveys were completed anonymously. The data collected formed a foundation for future decision-making for the author's program as to whether to continue incorporating Q-interactive into the assessment course. By sharing the findings from the data collected, other trainers can be helped with decisions related to

this technology inclusion, especially now as trainers are forced to teach in new and innovative ways. Additionally, a call can be made for additional practical research studies on this and other new technologies in our field for the mutual benefit of both trainers, students, and future clients our students will serve.

Participants and Procedure

The participants were graduate students ($n=46$) enrolled in the required cognitive assessment course. The student data was collected over three semesters (semester 1 = 18 students, semester 2=17 students, semester 3=11 students). Each graduate student received explicit instruction in the traditional paper format of the WISC-V and then in Q-interactive. They also participated in online tutorials, independent practice, and readings. Then, each student was randomly placed into an administration format group (traditional or digital) and observed either administering a traditional WISC-V or observed administering the iPad-mediated Q-interactive. Observations were conducted through GoReact where only the examiner was visible. Using the *Errors Checklist* (adapted from Oak et al., 2019), each administration was scored by the instructor. The same instructor did all the scoring. Errors from each administration were recorded onto a de-identified database of error information. Records of errors were kept by the course instructor. Then, the format groups switched and were observed and graded administering the opposite format of the WISC-V (the second observation results are not shared in this article).

Additionally, each of the graduate students enrolled in the cognitive assessment course were sent via email an electronic survey that would rate their perceptions of the traditional vs digital formats and training process. The survey was developed by the author, and the development was guided by the Checklist for Reporting Results of Internet E-Surveys (Eysenbach, 2004). The survey was field-tested by 5 practicing school psychologists for clarity, content, and usability. Items were amended, deleted, or added in accordance with the feedback received. The survey was hosted on the Qualtrics platform. Survey participation was voluntary, and all participants were provided with an electronic informed consent document. The survey was completed online, and no limitations were placed on how participants accessed the survey (i.e., they could gain access via computer, tablet, or other device). Unique IP addresses were required so that no single person could complete the survey multiple times. The survey consisted of 20 questions, of which 15 were multiple-choice items and 5 open-ended. The surveys were disseminated the last week of each of the semester, and one reminder email was sent the day before the survey closed on the final day of the semester. The survey email contained an introduction and explanation to the survey, which served as informed consent. All data were collected anonymously. The response

rate was 43%. Participants were not informed of the results of the survey.

Errors Checklist

The *Errors Checklist* used in this study was derived from existing skill evaluation checklists used in psychoeducational assessment courses, and it adapted aspects of other administration and scoring checklists utilized in the literature (Oak et al., 2019). The *Error Checklist* delineated possible errors for each subtest that were broken down into three categories, administration errors, scoring errors, and recording errors (Oak et al., 2019). Administration errors were defined as incorrect implementation of any of the various test administration rules that are in the WISC-V administration manual. Examples of administration errors included incorrect start points, not establishing basals correctly, discontinuing incorrectly, and failing to query correctly. Administration errors also included incorrect presentation of materials, not prompting when required, or not pointing at visual items when prompted. Scoring errors were defined as incorrectly scoring an examinee's response to an item, incorrectly adding individual scores to obtain a total raw score, and incorrectly converting a raw score to a scaled score. Recording errors were defined as errors resulting from a failure to record specific information or to record verbatim responses. The frequency and type of each individual error was recorded.

Results

A total of 46 WISC-V protocols and corresponding video recordings were obtained, 23 traditional paper and pencil, and 23 Q-interactive digital protocols, from 46 graduate student participants. The 46 WISC-V protocols and corresponding video recordings were rated to determine the frequency of administration, scoring, and recording errors between methods. Then, all 46 participants were surveyed about their perceptions of the two administration formats.

Traditional Format Errors vs Digital Format Errors

The results of this study indicate that the graduate student participants made a total of 79 errors across 23 protocols on the traditional format. A total of 15 errors were made across 23 digital protocols. Table 1 summarizes the frequency of errors by type and format.

The traditional format had an overall error rate of 100%. Every traditional administration had at least one error, whereas the digital format administrations had an overall error rate of 40%, with just 9 of the digital administrations having errors.

Table 1 Frequency (number and percentage) of errors by type and format

Error type	n=23 Trad format		n=23 Digital format	
	NE	%	NE	%
	Administration	57	72	9
Scoring	21	26	6	40
Recording	1	2	1	1

NE number of errors

Table 2 presents the frequency and rank order of specific errors by format.

Although the digital format produced less errors, the errors that did occur were on subtests that require clinical acumen. Consistent with Clark et al. (2017) previous study, Q-interactive did not reduce errors related to more complex judgments such as the nuanced scoring of the similarities and vocabulary subtests.

Survey

Table 3 summarizes graduate student’s perceptions of the two administration formats. Only 43% of the 46 graduate students responded to the optional survey (n=20).

Large Majority Preference for the Digital Format When asked an open-ended question about why they preferred the digital format, the 89% of the respondents who preferred the digital

format answered that “Q-interactive is very easy to use, and I feel there is less room for errors,” “my client was more engaged with Q-interactive,” and “the Q-Interactive is easy to administer, and I could focus on my student more than worrying about aspects of the test.” Other responses included “I liked that everything was right in front on one screen including the timers,” “Using the Q-interactive allowed better use of time as well as eliminated human error,” “It is clear and it automatically adjusts for basal and ceiling,” and “The ease of accessing important information right on the screen. The fact that it reverses when needed and the volunteer doesn’t even realize it where with the standard format the volunteer sees you flipping backward.” “Everything is contained in the i-pads for the administration; you don’t have to have wi-fi once you have downloaded the test; much easier to score and get the results; seemed to engage my volunteer more; don’t have to lug the suitcase of materials around.”

The remainder of the survey questions focused on each participant’s experience and perceptions of the volunteer taking their test. A large majority, 94%, of the respondents felt that their volunteer was eager to participate when presented with the iPads in the digital format, whereas only 29% felt that their volunteer was eager to participate when presented with the standard materials. A complete 100% of the respondents felt that their volunteer was engaged by the tasks and materials of the digital format; yet only 64% felt that their volunteer was engaged by the tasks and materials of the traditional format. Regarding enjoyment of the testing experience, 93% of the respondents felt that their volunteer enjoyed their experience with the digital format, while only 58% of the respondents felt that their volunteer enjoyed their

Table 2 Frequency and rank order of specific errors by format

Error description	Subtest where error occurred	Error type	Trad		Digital	
			%	RO	%	RO
Failure to administer 1 digit per second	Digit span	A	20	1	0	0
Failure to query the examinee when instructed by the manual	Similarities Vocabulary	A	19	2	53	1
Assigning an incorrect point value to a response	Similarities Vocabulary	S	16	3	40	2
Failure to expose the stimulus page for correct time	Picture span	A	12	4	0	0
Failure to read directions verbatim	Symbol search Digit span	A	11	5	3	3
Incorrect calculation of raw scores	Vocabulary	S	10	6	0	0
Failure to administer sample/practice/teaching items	Block design	A	4	7	0	0
Failure to place stimulus book/iPad properly	Block design	A	3	8	3	4
Failure to establish a basal	Similarities	A	3	9	0	0
Failure to record responses verbatim	Vocabulary	R	2	10	1	5

Error type A, administrative error; error type S, scoring error; error type R, recording error

Table 3 Graduate student's perceptions of traditional vs digital administration formats

	Trad %	Trad <i>n</i>	Digital %	Digital <i>n</i>
Administration was easy				
Strongly agree	25	5	70	14
Somewhat agree	45	9	25	5
Neither agree nor disagree	15	3	0	0
Somewhat disagree	10	2	5	1
Strongly disagree	0	0	0	0
Scoring was easy				
Strongly agree	20	4	85	17
Somewhat agree	50	10	10	2
Neither agree nor disagree	5	1	5	1
Somewhat disagree	20	4	0	0
Strongly disagree	5	1	0	0
My volunteer seemed engaged by the tasks and materials				
Strongly agree	25	5	85	17
Somewhat agree	40	8	15	3
Neither agree nor disagree	5	1	0	0
Somewhat disagree	30	6	0	0
Strongly disagree	0	0	0	0
My volunteer appeared eager to participate when presented with the testing materials				
Strongly agree	20	4	90	18
Somewhat agree	55	11	10	2
Neither agree nor disagree	15	3	0	0
Somewhat disagree	10	2	0	0
Strongly disagree	0	0	0	0
My volunteer appeared to enjoy their testing experience				
Strongly agree	25	5	80	16
Somewhat agree	35	7	15	3
Neither agree nor disagree	35	7	5	1
Somewhat disagree	0	0	0	0
Strongly disagree	5	1	0	0
Format I prefer				
Strongly agree	0	0	85	17
Somewhat agree	15	3	5	1
Neither agree nor disagree	0	0	0	0
Somewhat disagree	30	6	10	2
Strongly disagree	55	11	0	0

experience with the standard format. Respondents reported liking the tablets because they are easier to transport than traditional test kits. Finally, 88% of the survey respondents indicated that they preferred learning the traditional paper format first, while 11% indicated that they felt learning the Q-interactive first would be more

beneficial. Pearson (2017) recommends that Q-interactive be taught in conjunction with traditional paper administration.

Discussion

Coyne and Bartram (2006) stated that the advances in technology related to psychological assessment require a re-examination of the training of those who use tests and that training in this area is lagging and there is an urgent need to catch-up. This is even more true today as we face technology- and assessment-related issues during the COVID-19 pandemic. The current study sought to investigate Q-interactive as an instructional tool used within an assessment course. Specifically, does Q-interactive alleviate some of the administration and scoring errors that are common to students learning these tests? And what are graduate student's perceptions of traditional vs digital assessment formats? Results revealed several key findings.

Error Rates First, administrations of the WISC-V using traditional format with newly trained graduate students had higher amount of errors than the errors made using the digital format. As computation errors typically compose over one-third of the total errors made on graduate student protocols (Loe et al., 2007), the digital format reduced clerical and procedural errors. This finding regarding error rates was not totally unexpected and similar to previously conducted studies, where errors were found on a majority of the traditional graded protocols (for instance, Loe et al. (2007) found 98% and Belk et al. (2002) found 100% had errors). Because teaching these procedures and addressing these types of errors takes a good deal of instructional time when teaching traditional Wechsler test administration, valuable instructional time can be freed for more in-depth treatment of other issues, such as rapport building, the development of clinical acumen, score interpretation, and case conceptualization. Regardless of the administration format, a thorough understanding of scoring is fundamental to the competent administration and interpretation of assessment data.

Perception of Graduate Students Second, results revealed that 89% of the graduate students preferred the digital format when it was taught to them after the traditional format, so they knew both formats and had a chance to try them back-to-back. While the digital format has some reported shortcomings, the strength of its higher accuracy rate and higher positive rating by the user makes the Q-interactive an approach that is hard to ignore, given the current pandemic crisis and new trends in e-learning and online delivery systems. Additionally, with Q-interactive

gaining popularity with school districts and the need for alternative modes for assessment in school districts facing COVID-19 limitations and challenges, trainers must adapt adequately to prepare their students for effective school-based practice.

Implications

The current study has noteworthy implications for trainers of future school psychologists who teach assessment courses. Although the field of school psychology will most likely continue to move toward using technology to administer test batteries (Dumont et al., 2014), teaching clinical, observational, and interpretation skills is a critical component of the assessment courses, regardless of whether it is digital or traditional format. With the error rates being lower with the digital assessment, trainers can focus on developing those higher level skills in their candidates rather than focusing so heavily on test administration.

An unforeseen benefit of incorporating Q-interactive into recent assessment courses was that it prepared students for internships and new employment during COVID-19. This unprecedented, “new era of assessment” within a sea of change and uncertainty was a little more manageable as they became leaders in school districts implementing Q-interactive as a safe alternative for testing that can be conducted 6 feet apart. Graduate students must have awareness of the future of psychological testing and be prepared to participate effectively in the field as it moves forward (Gabel, 2013).

Limitations

As with any study, there are limitations to the current investigation. First, participants were not asked demographic questions that could have made them identifiable. The hope was that as much anonymity as possible improved the likelihood that they would participate and complete the survey. Thus, comparisons between groups of participants were not possible. Second, the questionnaire itself was developed by the author specifically for the purposes of obtaining feedback from graduate students. The questions developed have good face validity for this purpose, but other measures of validity were not calculated. Tertiary, the size of the sample was small and limited, dictated by course enrollment during the three consecutive semesters of this study. Although every effort was made to randomize administration format group assignment, the final groupings were determined by semester of course enrollment and thus adversely affects

generalizability. Additionally, the survey response rate was poor with only 20 out of 46 students participating in the survey (46% response rate). Also, time invested by the graduate students for independent practice of Q-interactive versus traditional assessment was not measured.

Further Research and Conclusion

The results of this study provide valuable information regarding Q-interactive as an instructional tool for graduate education. A more intensive evaluation of the digital format is called for with graduate students as well as continuing education for professionals, since the data show practitioners do not reduce error rates (and may, indeed, fossilize them) over time when using traditional models. Additionally, studies that investigate the psychometric equivalency of Q-interactive and remote assessment practices should be completed knowing that the instruments used by school psychologists (cognitive ability/intelligence tests, processing tests, neuropsychological tests, achievement tests, etc.) have not been normed or validated to be used in remote administrations (Hiramoto, 2020). Technology promises many benefits, yet we must continue to embrace evidence-based solutions to the challenges we face. Given the critical decisions—in education, special education eligibility, prisons, and other fields—made on the bases of individual intelligence tests, few topics deserve our more immediate attention.

Declarations

Research Involving Human Participants and/or Animals All procedures conducted in this study involving human participants were in compliance with the ethical standards of the institutional and/or national research committee, and IRB approval for the study was provided prior to any collection of data.

Informed consent Informed consent was obtained from all individual participants included in the study.

Competing interests The authors declare no competing interests.

References

- Alfonso, V. C., Johnson, A., Patinella, L., & Rader, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. *Psychology in the Schools, 35*, 119–125.
- Alfonso, V. C., LaRocca, R., Oakland, T. D., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review, 29*, 52–64.
- Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment, 7*, 119–129.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bauer, R. M., Iverson, G. L., Cemich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology*, *27*(3), 362–373.
- Belk, M. S., LoBello, S. G., Ray, G. E., & Zachar, P. (2002). WISC-III administration, clerical, and scoring errors made by student examiners. *Journal of Psychoeducational Assessment*, *20*, 290–300.
- Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology*, *72*, 29–48.
- Butcher, J. N. (2003). Computerized psychological assessment. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: assessment psychology* (pp. 141–163). Hoboken: John Wiley & Sons.
- Castillo, J., Curtis, M., & Gelley, C. (2012). Professional practice school psychology 2010—part 2: School psychologists' professional practices and implications for the field. *Communiqué*, *40*, 4–6.
- Cayton, T., Wahlstrom, D., & Daniel, M. (2012). The initial digital adaptation of the WAIS-IV. In E. Lichtenberger & A. Kaufman (Eds.), *Essentials of WAIS-IV assessment* (pp. 389–427). Hoboken: John Wiley.
- Charter, R. A., Walden, D. K., & Padilla, S. P. (2000). Too many simple clerical scoring errors: The Rey Figure as an example. *Journal of Clinical Psychology*, *56*, 571–574.
- Clark, S. W., Gulin, S. L., Heller, M. B., & Vrana, S. R. (2017). Graduate training implications of the Q-interactive platform for administering Wechsler intelligence tests. *Training and Education in Professional Psychology*, *11*(3), 148–155.
- Coyne, I., & Bartram, D. (2006). Design and development of the ITC guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*(2), 133–142.
- Curtis, M. J., Hunley, S. A., & Grier, J. E. (2002). Relationships among the professional practices and demographic characteristics of school psychologists. *School Psychology Review*, *31*, 30–42.
- Dumont, R., Viesel, K. D., Kohlhagenis, J., & Tabibis, S. (2014). A review of Q-interactive assessment technology. *Communiqué*, *43*(1), 8–12.
- Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, *6*(3), e34. <https://doi.org/10.2196/jmir.6.6.e34>.
- Fagan, T. K., & Wise, P. S. (2007). *School Psychology: Past, present, and future* (3rd ed.). Maryland: National Association of School Psychologists.
- Farmer, R. L., McGill, R. J., Dombrowski, S. C., Benson, N. F., Smith-Kellen, S., Lockwood, A. B., Powell, S., Pynn, C., & Stinnett, T. A. (2020). Conducting psychoeducational assessments during the COVID-19 Crisis: The danger of good intentions. *Contemporary School Psychology*, *24*, 1–6. <https://doi.org/10.1007/s40688-020-00293-x>.
- Foxcroft, C., & Davies, C. (2006). Taking ownership of the ITC's guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*(2), 173–180.
- Hosp, J. L., & Reschly, D. J. (2002). Regional differences in school psychology practice. *School Psychology Review*, *31*, 11–29.
- Gabel, A. (2013). Teaching assessment: Evolved- Session 3. <http://www.pearsonassessment.com>
- Gilbert, K., Kranzler, J. H., & Benson, N. F. (2020). *Effect of WISC-V type of administration on test performance*. Baltimore: Poster presented at the meeting of the National Association of School Psychologists. MD.
- Gurley, J. R. (2008). An examination of scoring accuracy on intelligence and achievement measures (Doctoral dissertation). Retrieved from ProQuest. (Accession No. 3329506) <http://www.proquest.com/productservices/pqdt.html>
- Hiramoto, J. (2020). Mandated special education assessments during the COVID-19 shutdown (California Association of School Psychologists position paper). Retrieved from <http://www.casonline.org>.
- Krach, S. K., McCreery, M. P., Dennis, L., Guerard, J., & Harris, E. L. (2019). Independent evaluation of Q-interactive: A paper Equivalency comparison using the PPVT-4 with preschoolers. *Psychology in the Schools*, *57*(1), 17–30.
- Lockwood, A. B., & Farmer, R. L. (2019). The cognitive assessment course: Two decades later. *Psychology in the Schools*, *57*(2), 265–283.
- Loe, S. A., Kadlubek, R. M., & Marks, W. J. (2007). Administration and scoring errors on the WISC-IV among graduate student examiners. *Journal of Psychoeducational Assessment*, *25*(3), 237–247.
- McDermott, P. A., Watkins, M. W., & Rhoad, A. M. (2014). Whose IQ is it? Assessor bias variance in high-stakes psychological assessment. *Psychological Assessment*, *26*(1), 207–214.
- Mrazik, M., Janzen, T. M., Dombrowski, S. C., Barford, S. W., & Krawchuk, L. L. (2012). Administration and scoring errors of graduate students learning the WISC-IV: Issues and controversies. *Canadian Journal of School Psychology*, *27*, 279–290.
- National Association of School Psychologists. (2010a). *Model for comprehensive and integrated school psychological services*. Bethesda, MD: Author.
- National Association of School Psychologists. (2010b). *Standards for graduate preparation of school psychologists*. http://www.nasponline.org/standards/2010standards/1_Graduate_Preparation.pdf
- Noland, R. M. (2017). Intelligence testing Using a tablet computer: Experiences with using Q-interactive. *Training and Education in Professional Psychology*, *11*(3), 156–163.
- Oak, E., Viesel, K. D., Dumont, R., & Willis, J. (2019). Wechsler administration and scoring errors made by graduate students and school psychologists. *Journal of Psychoeducational Assessment*, *37*(6), 679–691.
- Oakland, T. D., & Jimerson, S. R. (2006). School psychology: A retrospective view and influential conditions. In S. R. Jimerson, T. D. Oakland, & P. Farrell (Eds.), *The handbook of international school psychology* (pp. 453–462). Thousand Oaks: Sage Publications.
- Ramos, E., & Alfonso, V. (2009). Graduate Students' Administration and scoring errors on the Woodcock-Johnson III Tests of Cognitive Abilities. *Psychology in the Schools*, *46*(7), 650–657.
- Ready, R. E., & Veague, H. B. (2014). Training in psychological assessment: Current practices of clinical psychology programs. *Professional Psychology: Research and Practice*, *45*(4), 278–282.
- Scheller, A. (2013). Q-interactive: Overview. <http://www.pearsonassessment.com>
- Schulenberg, S. E., & Yutzenka, B. A. (2004). Ethical issues in the use of computerized assessment. *Computers in Human Behavior*, *20*(4), 477–490.
- Simons, R., Goddard, R., & Patton, W. (2002). Hand-scoring error rates in psychological testing. *Assessment*, *9*(3), 292–300.
- Slate, J. R., & Chick, D. (1989). WISC-R examiner errors: Cause for concern. *Psychology in the Schools*, *26*, 78–83.
- Slate, J. R., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale-Revised: An empirical evaluation of practice administrations. *Professional Psychology: Research and Practice*, *22*(5), 375–379.
- Slate, J. R., Jones, C. H., Murray, R. A., & Coulter, C. (1993). Evidence that practitioners err in administering and scoring the WAIS-R.

- Measurement and Evaluation in Counseling and Development*, 25(4), 156–161.
- Snyder, D. K. (2000). Computer-assisted judgment: Defining strengths and liabilities. *Psychological Assessment*, 12(1), 52–60.
- Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools*, 51(10), 1031–1045.
- Styck, K. M., & Walsh, S. M. (2016). Evaluating the prevalence and impact of examiner errors on the Wechsler scales of intelligence: A meta-analysis. *Psychological Assessment*, 28(1), 3–17. <https://doi.org/10.1037/pas0000157>.
- U.S. Department of Education. (2020). *Implementation of IDEA Part B Provision of Services in the COVID-19 environment*. Office of Special Education and Rehabilitative Services. <https://www2.ed.gov/policy/speced/guid/idea/memosdcltrs/qa-provision-of-services-idea-part-b-09-28-2020.pdf>
- Weiner, I., & Greene, R. (2007). *Handbook of personality assessment*. New York: John Wiley.
- Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, 25(1), 9–23.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Stephanie Corcoran is a nationally certified school psychologist and currently serves as an assistant professor at the University of Alabama at Birmingham (UAB) where she directs the School Psychometry Program and serves an associate scientist for the Civitan International Research Center.