Equivalence of Remote, Online Administration and Traditional, Face-to-Face Administration of the *Reynolds Intellectual Assessment Scales-Second Edition*

A. Jordan Wright, Ph.D., ABAP
Steinhardt School of Culture, Education, and Human Development
New York University

January, 2018

Introduction

Traditional, in-person testing has been at the core of psychological assessment, but today's technological advances are inspiring new avenues for testing. Remote, online administration of assessments can potentially alleviate issues that many educational organizations are now facing. These include localized staffing shortages, an increasing volume of referrals in general, and seasonal spikes in referrals, all which make it a challenge to complete assessments within a timely manner. These demands mean school psychologists may not be adequately providing preventative services, like addressing student behavior, mental health concerns, and school climate issues, simply for lack of time.

PresenceLearning, Inc. has been providing "teletherapy assessment" in the speech and language and occupational therapy realms since 2009. While online rating scale types of assessment have become very prevalent, individually administered psychoeducational assessments online are a relatively new concept. Individual testing as part of psychoeducational assessment is often a time-intensive endeavor that can be further complicated by scheduling conflicts. By reducing the need to travel and potentially increasing the number of providers available at different times, remote testing is a convenient resource to increase access to assessment services.

PAR, Inc. and PresenceLearning, Inc. have collaborated to develop a process for delivering online administration of the *Reynolds Intelligence Assessment Scales-Second Edition* (RIAS-2; Reynolds & Kamphaus, 2015). This study evaluates the equivalence between remote, online administration and traditional, in-person administration of the RIAS-2 for children between the ages of 3 and 19. The goal is to evaluate the scatter of scores captured by both administration formats (with randomly assigned participants), in order to explore if the formats are interchangeable.

The RIAS-2 includes four core subtests: Guess What, Odd-Item Out, Verbal Reasoning, and What's Missing. These four core subtests combine to form the Composite Intelligence Index, which is an estimate of global intelligence. There are two sets of supplemental subtests as well. The Verbal Memory and Nonverbal Memory subtests combine for a supplemental Composite Memory Index (CMX). The Speeded Naming and Speeded Picture Search subtests combine to form a Speeded Processing Index (SPI).

There are many factors involved with administration of any assessment that can impact the examinee's performance, such as room conditions and outside distractions. When converting traditional, in-person test administration to online, remote administration, the alteration from physical stimulus books to a digital format and the dynamic of testing with an examiner through telecommunication may also impact the examinee's responses. Therefore, it is important to examine whether or not the results from the RIAS-2 are generally equivalent in these two different administration formats. The study discussed here presents the psychometric equivalence results between these two testing methods.

Fidelity Requirements

The online remote administration RIAS-2 data were collected under very specific conditions, which are outlined below. It should be noted that the results of the present study should only generalize to testing situations that conform to the following fidelity requirements. Other remote, online administrations of the test have not been evaluated for their potential equivalence to traditional, in-person administration of the test.

Physical Conditions

The online, remote administration of the RIAS-2 must take place in a quiet room where the examinee is seated in front of a computer (with a monitor at least 15 inches in size), through which he or she verbally and visually communicates with the examiner. Two high-definition cameras are set up so that the examinee's face (one camera) and desk/work space (the other camera) are visible, while the examinee can see the examiner's face and digital materials via the digital platform on the monitor. The examinee also has a headset with a microphone and a mouse to indicate response choices on the screen. The proctor, who remains in the room with the examinee, assists the examiner by manipulating materials and redirecting the examinee as directed by the examiner.

Digital Platform

The RIAS-2 stimulus books were converted to digital images on the PresenceLearning platform. Careful consideration was given to the fidelity of the figures on all subtests. In order to maintain accurate presentation of the stimuli, the equipment used must have at least 800 x 600 pixels of resolution quality on both monitors (the examiner end and the examinee/test-taker end). The audio was presented through the platform to ensure clarity and quality (rather than through other means, like a conference phone). The examinee's screen was at least 15 inches in size, and the examinee utilized a separate mouse and headset. Additionally, the proctors (sitting with the examinees) had their own headset.

For this study, the platform required the examiner to log into the PresenceLearning platform with a password, and then the examiner had to "admit" the examinee into the test administration. In other words, having the URL did not allow the examinee to log in and view the test materials without the examiner explicitly allowing access in order to maintain test security.
After being given instruction guides on how to install the headset and document camera equipment, each examiner completed a technology check with PresenceLearning client services and support staff.

Examiner Training

All of the examiners for this study were school psychologists or clinical psychologists with previous knowledge and experience administering the RIAS-2 in the traditional, in-person format. Some examiners also had previous experience using the PresenceLearning platform. Examiners not familiar with the PresenceLearning platform underwent additional, specific training provided by PresenceLearning staff. This consisted of a step-by-step training video to

familiarize the examiner with the PresenceLearning platform, a group session to guide examiners through RIAS-2-specific training, and individual practice and feedback sessions as needed.

Proctor Training

The proctors were recruited and trained by the examiners. Proctors were first asked to review a PresenceLearning introductory document online. Proctors were then coached by their examiners on when they were and were not allowed to speak to an examinee.

Equivalence Study Design

To reduce confounding factors, this study utilized a case-control match design format. Examinees were paired by age and gender and were randomly assigned to take either the traditional, in-person RIAS-2 or the remote, online RIAS-2. The two samples (in-person and remote) were equal in number and should be generally equivalent on potentially confounding variables and general cognitive ability (because of the random assignment).

For the purposes of this study, both significance tests (p values of t-tests) and effect sizes (omega squared, because of the limited sample size) were calculated to determine equivalence. The standards of $p \geq .05$ and $\omega^2 \leq .03$ were selected as cut-off criteria for a significant effect of the administration type. Omega squared (ω^2) was selected based on the design and power with the sample size (see Olejnik & Algina, 2003), and the criterion cut-off was selected to correct for power issues with a small sample size (see Cohen, 1992 and Button et al., 2013).

RIAS-2 Equivalence Study

Participants

PAR, Inc. utilized known examiners to recruit a sample of 104 examinees between the ages of 3 and 19. Each of these examinees was matched with a 'pair'—a person of the same gender, age, race, education level, and geographical region (see Table 1). Examiners were asked to recruit twin pairs when possible; the final sample included 14 pairs of twins. Payment was provided to all examinees.

Demographic characteristics of the sample are presented in Table 1. There was an overall equal representation of males and females. In terms of race and ethnicity, the current sample closely resembled current (2016) U.S. census proportions, with Caucasians being very slightly overrepresented (versus the census data of 61%), and African American, Hispanic, and Other races/ethnicities being represented within 3% of the census.

Table 1
Demographic Characteristics of the Sample

Demographic Characteristic		Administration Format			
		Traditional, In-Person		Remote, Online	
		Number of Cases			
		52		52	
		Male	Female	Male	Female
Age (years)	3-5	5	4	5	4
	6-9	6	9	6	9
	10-14	7	6	7	6
	15-19	8	7	8	7
	Mean	10.5		10.5	
	SD	4.72		4.72	
Race/Ethnicity	White	63%		62%	
	Black	12%		12%	
	Hispanic	19%		21%	
	Other	6%		6%	
Parent Education	Less than HS graduate	10%		10%	
	HS graduate	10%		8%	
	Some College	33%		33%	
	College Graduate	48%		50%	

Procedure

Data were collected between July and September 2017 using eight examiners who tested examinees in Arkansas, Georgia, Idaho, Kansas, Mississippi, New York, Ohio, Tennessee, and Texas. Examiners and proctors were paid per completed case.

All RIAS-2 test administrations, regardless of format/condition, occurred within the examinees' school, in the examiner's office, or in the examinee's home, and all examiners administered the test in both formats. The RIAS-2 was administered in both formats according to the standardized procedures prescribed by PAR, Inc.

Results

The means and standard deviations of all RIAS-2 subtest T scores and index standard scores for each administration format and for the total sample are presented in Table 2.

Table 2
Descriptive Statistics for the Test Scores by Administration Format

Subtest/Index	Traditional, In-Person Administration		Remote, Online Administration		Total Sample	
	Mean	SD	Mean	SD	Mean	SD
Subtest Scores						
Guess What (GWH)	49.98	11.88	47.10	11.06	48.54	11.51
Odd-Item Out (OIO)	52.65	10.49	52.06	10.23	52.36	10.31
Verbal Reasoning (VRZ)	49.62	11.99	47.62	10.77	48.62	11.39
What's Missing (WHM)	56.12	9.33	52.40	12.46	54.26	11.11
Verbal Memory (VRM)	47.19	10.32	47.04	11.28	47.12	10.76
Nonverbal Memory (NVM)	50.88	7.61	53.44	7.14	52.16	7.45
Speeded Naming Task (SNT)	52.31	9.28	48.44	11.55	50.38	10.61
Speeded Picture Search (SPS)	47.81	9.91	43.81	13.91	45.81	12.18
Index Scores						
Verbal Intelligence Index (VIX)	99.56	16.96	95.94	15.39	97.95	16.22
Nonverbal Intelligence Index (NIX)	106.31	12.86	102.94	12.85	104.63	12.90
Composite Intelligence Index (CIX)	103.27	15.02	99.25	13.56	101.26	14.38
Composite Memory Index (CMX)	98.08	14.17	100.38	14.54	99.23	14.33
Speeded Processing Index (SPI)	100.25	14.70	92.88	19.82	96.57	17.75
N	52		52		104	

Note. All subtest scores are *T* scores ($M = 50$, $SD = 10$). All index scores are standard scores ($M = 100$, $SD = 15$).

Table 3 shows the comparisons between administration formats, both with hypothesis testing (*T* and *p* values) and effect sizes (ω^2). While most subtests and indices fell below the threshold of significance ($p < .05$ and $\omega^2 > .03$), the Speeded Processing Index (SPI) showed a significant effect for administration format, such that scores from the traditional, in-person format were significantly higher than scores from the online, remote format.

Table 3

Significance and Effect Size of Remote, Online Format on the Subtest and Index Scores of the RIAS-2

Subtest/Index	T	p	Effect Size
Subtest			
Guess What (GWH)	1.282	.203	.006
Odd-Item Out (OIO)	0.293	.770	-.009
Verbal Reasoning (VRZ)	0.895	.373	-.002
What's Missing (WHM)	1.720	.089	.018
Verbal Memory (VRM)	0.073	.942	-.010
Nonverbal Memory (NVM)	-1.768	.080	.020
Speeded Naming Task (SNT)	1.881	.063	.024
Speeded Picture Search (SPS)	1.689	.094	.018
Index			
Verbal Intelligence Index (VIX)	1.138	.258	.003
Nonverbal Intelligence Index (NIX)	1.335	.185	.007
Composite Intelligence Index (CIX)	1.432	.155	.010
Composite Memory Index (CMX)	-0.820	.414	-.003
Speeded Processing Index (SPI)	2.152	.034	.034

Note. A positive effect size indicates higher scores with traditional, in-person administration.

While one of the speeded tasks (Speeded Naming) was administered via the platform in the remote, online administration, the other, Speeded Picture Search, was actually administered using the same physical materials (not via the digital platform), so differences in performance attributable to the format was hypothesized to possibly be an artifact of some distraction due to having the platform visible during completion of the tasks. Research has suggested that voluntary attention improves developmentally with age (e.g., Enns & Brodeur). As such, data were analyzed for children aged 7 and higher for these tasks, to see if the significant differences by format remained. Table 4 presents means and standard deviations, and Table 5 presents the significance tests and effect sizes for this reduced sample. As revealed in these tables, when younger children are taken out of the sample, for children aged 7 and older, there was no effect for the different methods on performance.

Table 4

Descriptive Statistics for the Test Scores by Administration Format for Children 7-19

Subtest/Index	Traditional, In-Person Administration		Remote, Online Administration		Total Sample	
	Mean	SD	Mean	SD	Mean	SD
Subtest Scores						
Speeded Naming Task (SNT)	52.83	9.56	49.50	11.63	51.16	10.71
Speeded Picture Search (SPS)	50.40	7.52	47.93	9.66	49.16	8.69
Index Scores						
Speeded Processing Index (SPI)	103.12	12.85	97.70	16.24	100.41	14.80
N	40		40		80	

Note. All subtest scores are *T* scores (*M* = 50, *SD* = 10). All index scores are standard scores (*M* = 100, *SD* = 15).

Table 5

Significance and Effect Size of Remote, Online Format on the Speeded Processing Subtest and Index Scores of the RIAS2 for ages 7+

Subtest/Index	*T*	*p*	Effect Size
Speeded Naming Task (SNT)	1.397	.166	.012
Speeded Picture Search (SPS)	1.279	.205	.008
Speeded Processing Index (SPI)	1.657	.102	.021

Note. A positive effect size indicates higher scores with traditional, in-person administration.

Discussion

The present study aimed to examine the equivalence between the traditional, in-person administration of the *Reynolds Intelligence Assessment Scales-Second Edition* (RIAS-2) and an online, remote administration of the test. For the four core RIAS-2 subtests (which constitute the CIX) and the memory subtests (which constitute the CMX), there was no significant effect for administration procedure. As such, the formats are generally equivalent, and the same norms can be used. Across all ages, for clinicians who only administer the core four subtests or those plus the memory subtests, the administration procedures can be used interchangeably.

For the speeded tasks, there was a significant method effect, such that students evaluated in the traditional format performed significantly better than those evaluated via the online, remote platform. The two administration methods are not equivalent across the age span. Upon further

analysis, this only held true for students under the age of 7. As such, for children under 7 years old, it is not recommended to use the speeded processing subtests in the remote, online format.

For those children aged 7 and older, the two administration methods did not exhibit significant differences for the speeded processing tasks. As such, like all the other subtests on the RIAS-2, for children aged 7 and older, the two administration procedures are comparable, and the original norms can be used.

The present study suggests that the test, when given in the remote, online format (*in the specified, faithful procedure specifically evaluated in this study*), is generally equivalent and can use the norms of the traditional test. However, it is recommended that for children under 7 years old, the speeded tasks (which are supplemental) not be administered in this format.

References

Button, K.S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Enns, J. T., & Brodeur, D. A. (1989). A developmental study of covert orienting to peripheral visual cues. *Journal of Experimental Child Psychology, 48*, 171–189.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales, Second Edition and the Reynolds Intellectual Screening Test* (2nd ed.). Lutz, LF: PAR, Inc.