Canadian Journal of School Psychology
27(4) 279–290

DOI: 10.1177/0829573512454106
http://cjs.sagepub.com

Administration and Scoring Errors of Graduate Students Learning the WISC-IV: Issues and Controversies

Martin Mrazik[1], Troy M. Janzen[1], Stefan C. Dombrowski[2], Sean W. Barford[1], and Lindsey L. Krawchuk[1]

Abstract

A total of 19 graduate students enrolled in a graduate course conducted 6 consecutive administrations of the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV, Canadian version). Test protocols were examined to obtain data describing the frequency of examiner errors, including administration and scoring errors. Results identified 511 errors on 94% of protocols with a mean of 4.48 errors per protocol. The most common errors were identified on the Vocabulary, Similarities, and Comprehension subtests, which comprised 80% of all errors. A repeated-measures ANOVA (analysis of variance) was not significant across six administrations, $F(5, 90) = 1.609$, $p = .166$, $\text{eta}^2 = .082$, although there was a trend in the data for a reduced number of errors with successive administrations. Results were consistent with other studies that have determined graduate student administration and scoring errors do not improve with repeated administrations. Implications and recommendations to reduce administration and scoring errors among graduate students were discussed.

Résumé

Un total de 19 étudiants des cycles supérieurs inscrits dans un cours de deuxième cycle ont mené 6 administrations consécutives de l'Échelle d'Intelligence de Wechsler pour Enfants, Quatrième édition (WISC-IV, version canadienne). Protocoles d'essai

[1]University of Alberta, Edmonton, Alberta, Canada
[2]Rider University, Lawrenceville, NJ, USA

Corresponding Author:
Martin Mrazik, 6-135 Education North, Department of Educational Psychology, University of Alberta, Edmonton, Alberta T6G 2G5, Canada
Email: mrazik@ualberta.ca

ont été examinés afin d'obtenir des données décrivant la fréquence des erreurs de l'examinateur incluant les erreurs de l'administration et de notation. Les résultats ont identifié 511 erreurs sur 94% des protocoles, avec une moyenne de 4.48 erreurs par protocole. Les erreurs les plus communes ont été identifiées aux sous-tests de Vocabulaire, Similitudes et Compréhension qui composaient 80% de toutes les erreurs. Une ANOVA à Mesures Répétées n'était pas significative à travers les 6 administrations, $F(5, 90) = 1.609$, $p = .166$, $eta^2 = .082$, bien qu'il y avait une tendance dans les données pour un nombre réduit d'erreurs avec administrations successives. Les résultats étaient constants avec d'autres études qui ont déterminé l'administration des étudiants des cycles supérieurs et les erreurs de notation ne s'améliorent pas avec administrations répétées. Implications et recommandations pour réduire les erreurs de l'administration et la notation parmi les étudiants des cycles supérieurs ont été discutés.

Keywords

WISC-IV, intelligence test instruction, examiner error, test reliability, psychological assessment

The Wechsler Intelligence Scale for Children, 4th edition (WISC-IV; Wechsler, 2003) is one of the most widely used intelligence scales in the world. As a Level C instrument, it requires advanced graduate training in psychology to learn its administration, scoring, and interpretation properties. The task of learning how to properly administer and score the WISC-IV is a challenging one. The administration manual provides in-depth instructions to guide the clinical examiner, but how this is taught in a classroom setting is left up to professors and to the students themselves. There are no specific training models provided by the WISC-IV manual that guide professors on teaching pedagogy. The belief that "practice makes perfect" (Loe, Kadlubek, & Marks, 2007) appears to be the default process used to guide learning. As a result, understanding common errors made by graduate students as they learn the WISC-IV plays a pivotal role in enhancing student learning.

The WISC-IV offers unique challenges to those learning its administration and scoring as it builds upon previous editions of the test but has also undergone considerable changes since its early editions. Select subtests were removed altogether (Picture Arrangement), and others were added (Picture Concepts). Most subtests were updated with new items and extended ranges of test items. Although the changes to the WISC-IV have helped to maintain and improve its psychometric properties, the essence of obtaining valid and reliable results remains with the examiner and the accuracy of their administration and scoring. Some subtests (Matrix Reasoning, Picture Concepts) have a narrow range of responses, making it easier for the examiner to evaluate the correctness of an examinee's answers. However, other subtests (Similarities, Vocabulary, Comprehension) require much more clinical

judgment, and it is these subtests that require the greatest degree of learning on the part of the examiner. It's not surprising that verbal subtests, including Vocabulary, Similarities, and Comprehension, have historically had the most errors for students learning the WISC (Belk, Lobello, Ray, & Zachar, 2002; Miller & Chansky, 1972; Slate & Jones, 1990).

It is expected that graduate students will make mistakes as they learn the WISC-IV (Alfonso, Johnson, Patinella, & Rader, 1998). Research on earlier versions of the WISC identified that graduate errors are common. For instance, Slate and Chick (1989) found that students made, on average, 15 errors per protocol on the WISC-R, with approximately two thirds of their sample participants making mistakes. However, previous research has also suggested that graduate students did not show a significant decrease in the number of errors over 5 to 10 practice administrations. Similarly, Belk et al. (2002) found a mean of 10.9 errors per protocol, and errors count did not significantly decrease over repeated administrations with the WISC-III. The same concern was identified in a recent study by Loe et al. (2007) with the WISC-IV where students did not show statistically significant improvement on administration, computation, and recording errors, over three assessments. Yet the underlying belief is that students will improve with experience and practice. Authors of previously published studies have provided recommendations for optimal methods of instruction. For instance, Alfonso et al. recommended that graduates undertake five to six administrations to ensure the examiner has successfully attained the appropriate skill level.

However, the issue of errors among those who administer the WISC is not just a concern for graduate students. Sherrets, Gard, and Langner (1979) selected 200 random samples of WISC-R protocols from a psychiatric facility and public schools. Clinicians included certified school psychologists and PhD-level clinical psychologists. The number of errors was reported to be alarmingly high, with approximately 50% of protocols having one error and 90% of examiners making at least one error. Bradley, Hanna, and Lucas (1980) contacted 63 members from the National Association of School Psychologists (NASP) and asked them to score two WISC-R protocols, an "easy" protocol and a constructed difficult protocol. Results indicated a wide range of results among the scoring, with errors and differences in clinical judgment of subjectively scored items contributing to the variation. Ryan, Prifitera, and Powers (1983) investigated the reliability between doctoral-level psychologists with graduate students scoring of two actual protocols and found that although there were no significant differences in scoring discrepancies on any subtest or IQ index means, the doctoral-level psychologists were more likely to have greater scoring variability on the Performance scale. To our knowledge, there have been no recently published studies with practicing psychologists used as participants.

In spite of the findings from the various studies of a high number of clerical, administration, and scoring errors, there has been little response to address these concerns from a teaching/pedagogical perspective. Slate and Chick (1989) argued that better training was needed to prepare graduate students and restated their beliefs a few years later, following an examination of WISC-III errors (Slate, Jones, & Murray, 1991,

1993). Other calls for better pedagogical methods have come from Belk et al. (2002), Loe et al. (2007), and Slate and Chick.

The purposes of this study were threefold. There are a number of potential sources for error, including addition/charting errors, recording error, and errors associated with graduate student learning. Thus, the first objective was to carefully identify and describe all sources of scoring errors by students as they learn the WISC-IV. The second objective was to evaluate graduate student learning by reviewing errors made over the course of six successive administrations of the WISC-IV. Finally, on the basis of results from this and other studies, we present a discussion and suggestions for more optimal training methods.

Method

A total of 21 students were enrolled in a full-year psychological testing course at an urban university in Western Canada. However, two students did not complete all six WISC assessments as they opted to administer another cognitive test (the Stanford Binet, 5th ed.) and were excluded from the study. Of the remaining 19 students, one student was a first-year PhD student and the remaining 18 students were second-year students enrolled in a master's-level program preparing them for professional practice in psychology. None of the students reported previous coursework or experience administering individually administered, Level C tests of intelligence. Each student completed a minimum of 6 WISC-IV assessments for a total of 114 assessments. Only the first six assessments were used in this study.

The data were collected during the course of the 2009-2010 academic year. Students attended class a total of 26 times, with each class lasting 2 hours and 50 minutes. The WISC-IV was taught over the course of three lectures and was the first intelligence test taught to students. Each student was provided with a WISC-IV kit for use during the academic year. All students were paired with a licensed psychologist in the local community who worked with the student for the entire year (primarily to assist with report writing). All supervising psychologists had at least 5 years of experience working as a registered psychologist in Alberta. Uniform instruction of the WISC was provided by the primary author to all students who met weekly for 3-hour classes. Whereas an entire year was devoted to the class, teaching of the WISC included four classes, including a brief history of the Wechsler scales, followed by detailed teaching of administration, scoring, and interpretation of the WISC-IV. Students completed two practice tests and were required to submit a sample video with an administration of the WISC-IV prior to working with clients. The graduate students' participants were children referred by their parents/guardians to the University of Alberta Education clinic, which is an outpatient community clinic. All referral participants and parents/guardians were aware that clinicians were graduate psychology students in training and provided written informed consent.

All students were required to submit WISC-IV protocols to one of two teaching assistants (TAs), both of whom were enrolled in a PhD program in psychology. Both TAs had completed the equivalent of three courses in psychological assessment. TAs were provided with a detailed rubric that contained every possible scoring error from

Table 1. Frequency of Administration and Recording Errors by Subtest on the Wechsler Intelligence Scale for Children (4th Ed.)

	Protocols with errors			Errors per protocol	
	N	%	Total errors	Mean	Range
Block design	33	27.5	40	0.33	0-2
Similarities (total errors)	57	48.0	115	0.96	0-9
Query errors only	32	26.7	52	0.43	0-6
Digit span	31	27.0	42	0.34	0-6
Picture concepts	4	3.0	4	0.03	0-1
Coding	1	1.0	1	0.01	0-1
Vocabulary (total errors)	63	59.0	144	1.20	0-6
Query errors only	43	36.0	71	0.59	0-6
Letter-number sequencing	9	8.0	9	0.08	0-1
Matrix reasoning	12	10.0	13	0.11	0-2
Comprehension (total errors)	73	61.0	153	1.28	0-14
Query errors only	47	39.0	82	0.68	0-9
Symbol search	5	4.0	5	0.04	0-1
Total errors	112	93.0	754	6.28	0-28

Note: *N* of protocols = 114.

the WISC-IV. This included a checklist of items from the front page of the WISC-IV (for instance, ensuring the client's age was calculated correctly, proper use of appropriate norm tables, addition of subscale scores, etc.) and then all the administration and scoring rules for each of the 10 core subtests as per the WISC-IV manual. Every test item on the Vocabulary, Similarities, and Comprehension subtests was evaluated for proper use of queries, prompts, and accuracy of assigned score (i.e., a score of 0, 1, or 2). Overall, there were 271 potential item errors reviewed for each WISC-IV protocol. Each student was given a grade for the accuracy of scoring, with each scoring error constituting a 0.5% deduction in their total score (a total score out of 100) for each report written by the student. Students were required to correct errors and resubmit them to the TAs for a second evaluation. Two practicing PhD-level psychologists (including the course instructor) separately reviewed 10% of the protocols to ensure interrater reliability. There was an agreement of 93% on the total errors noted.

Results

Results from this study indicated that graduate students committed scoring errors on the majority of test protocols (94%). This figure is similar to other recent studies (Loe et al., 2007) that found 98% of all protocols to have errors. There were a total of 511 errors with 107 of 114 protocols having errors, resulting in a mean of 4.48 errors per protocol (see Table 1). For the current study, there was a wide range of error totals

Table 2. Frequency of Errors for Similarities, Vocabulary, and Comprehension Subtests per Protocol

	Similarities		Vocabulary		Comprehension	
	N	%	*N*	%	*N*	%
Incorrect use of a query	31	26.5	30	34.2	71	60.7
Wrong starting point	1	0.9	0	0.0	0	0.0
Failure to use reversal rule for basal	5	4.3	4	3.4	3	2.6
Discontinued too early	1	0.9	0	0.0	0	0.0
Discontinued too late	5	4.3	2	1.7	1	0.9
Change in score	46	39.3	66	56.4	56	47.8
Failed to ask for second response	—	—	—	—	8	6.8
Failure to add items correctly	2	1.7	5	4.3	4	3.4

Note: *N* of protocols = 114

per protocol, ranging from no errors to 28 errors. Not surprisingly, subtests comprising the Verbal Comprehension Index had many more errors (406) than subtests comprising the Perceptual Reasoning Index (56). The Working Memory Index subtests had more errors (43) compared with the Processing Speed Index (6).

As anticipated, graduate students made the greatest number of errors on the Similarities, Vocabulary, and Comprehension subtests, which accounted for 80.2% of all errors (see Table 2). For the current study, any response where the query was missed, a query used incorrectly, or used when it shouldn't were identified as an error. Based on this procedure, it was found that problems with querying represented the most common source of student error on these three subtests (37.3% of all errors). Students had the most difficulty with querying on the Comprehension subtest. This is likely due to the wider range of possible responses given by test participants on this subtest. Identification of query errors led to changes in subtest scaled scores for a large number of the WISC-IV protocols. In contrast to query errors, graduate students made very few errors (5.9% of all protocols) due to inattention/human error (e.g., failure to add scores correctly, starting at the wrong item, etc.). The effects of errors on subtest and index scores were not calculated for each subject. However, based on the range of errors identified for a child of mean age from this study, the corrected VCI would range from 0 to 8 scale score points (mean = 3 points) higher, and the corrected FSIQ would range from 0 to 6 points (mean = 2 points) higher.

The front page of the WISC-IV provides the summary data for subtests and index scores and requires the use of appropriate norm tables located at the back of the manual. Completing this page requires careful attention to the age of the participant, adding subtest and index scores correctly, and using the appropriate normative tables. Results identified only a small proportion of students who made errors on the front page (4.3% of all errors) resulting in incorrect subtest and index score calculations (see Table 3). In several cases, the same student made errors on multiple occasions, indicating that some students were more inattentive and prone to making more human

Table 3. Frequency of Front-Page Errors on the Wechsler Intelligence Scale for Children (4th Ed.)

	Number of students making errors (total = 19)	
	N	Percent of all errors
Demographic information recorded incorrectly	1	6.0
Miscalculation of chronological age	1	6.0
Incorrectly transferred raw scores to front page	2	12.0
Failed to use correct norm tables	0	0
Failed to record raw scores for each subtests	4	24.0
Used wrong norm tables	4	24.0
Added Index/IQ scores incorrectly	2	12.0
Failed to fill in subtest profile	4	3.4
Failed to fill in composite score profile	4	3.4
Total errors	22	

Table 4. Errors by Administration for the Wechsler Intelligence Scale for Children (4th Ed.)

	Administration number					
	1	2	3	4	5	6
Block design	8	4	8	8	7	7
Similarities (total errors)	27	17	15	18	9	24
Digit span	6	5	7	4	11	6
Picture concepts	2	0	0	0	0	2
Coding	0	0	0	0	0	1
Vocabulary (total errors)	31	33	15	22	16	23
Letter-number sequencing	2	2	1	1	0	1
Matrix reasoning	2	2	1	3	2	3
Comprehension (total errors)	30	29	14	37	16	23
Symbol search	1	1	0	0	2	2
Total errors	109	93	61	93	63	92[a]

Note: *N* of protocols = 114

[a]Repeated-measures ANOVA (analysis of variance), $F(5, 90) = 1.609$, $p = .16$.

errors. As noted above, because students were penalized for any error, it is likely that students took careful precautions to ensure they didn't lose marks, which resulted in making very few errors due to inattention/carelessness.

Finally, a series of one-way repeated-measures ANOVAs (analyses of variance) was conducted to determine whether the number of errors for the entire administration, the Verbal Comprehension Index subtests, and then Vocabulary, Similarities, and Comprehension, changed over the course of six practice administrations. Table 4

describes mean error rates for each practice administration for the total administration, the Verbal Comprehension Index subtests, and then for Vocabulary, Similarities, and Comprehension. Over the six administrations, the repeated-measures ANOVA results were not significant for the total test, $F(5, 90) = 1.609$, $p = .166$, $\text{eta}^2 = .082$; the Verbal Comprehension Index, $F(5, 90) = 1.692$, $p = .145$, $\text{eta}^2 = .086$; Vocabulary, $F(5, 90) = 0.874$, $p = .502$, $\text{eta}^2 = .046$; Similarities, $F(5, 90) = 0.996$, $p = .425$, $\text{eta}^2 = .052$; and Comprehension, $F(5, 90) = 1.794$, $p = .122$, $\text{eta}^2 = .091$. Although the repeated-measures ANOVA was not significant for any of the analyses conducted, it is noteworthy that follow-up, pairwise comparisons between administrations revealed several significant effects for the total (1 with 3, $p < .01$; 2 with 3, $p < .05$; 3 with 6, $p < .05$) and the verbal index (1 with 3, $p < .01$; 2 with 3, $p < .0001$; 3 with 6, $p < .05$). A similar result was found for the comprehension subtest (1 with 3, $p < .001$; 3 with 4, $p < .05$; 3 with 6, $p < .001$) but not for the Vocabulary and Similarities subtests.

Discussion

Results from the current study continue to show that graduate students are prone to making multiple errors when administering and scoring the WISC-IV. Research on errors with previous versions of the WISC suggests a wide range of mean errors, from 7.8 errors per protocol to 45.2 per protocol. Our results were lower than most studies with a mean of 4.48 errors per protocol. However, findings were reasonably consistent with Loe et al.'s study who found a mean of 5.6 errors per protocol with the WISC-IV. Similar to previous studies, we found that errors were discovered on a majority of protocols (for instance, Loe et al., 2007, found 98% and Belk et. al found 100% had errors). Yet the decrease in mean errors per protocol identified compared with previous editions of the WISC is likely multifactorial. Primarily, the newest version of the WISC (WISC-IV) has improved scoring criterion, provided more defined outcomes for scoring on specific subtests, and new subtests altogether. For instance, Object Assembly (WISC-III) was a core subtest on earlier versions of the WISC, and there were more opportunities for making mistakes when scoring this subtest, compared with Picture Concepts (WISC-IV), which involves defined outcomes for scoring.

As identified by previous research, problems with querying responses for Similarities, Vocabulary, and Comprehension represent the most frequent and significant sources of error. In our study, errors with querying/prompting among subtests in the VCI represented 80% of all errors. Loe et al. (2007) also found a significant number of query errors but found more errors per protocol than our study. The evidence suggests that students clearly struggle with the concept of querying/prompting a response. This is likely a result of uncertainty of when to correctly query (whether missing a query or querying too much) and problems with conceptualizing what constitutes a 0, 1, or 2 point response. In spite of students being provided with corrective feedback in the form of being penalized for incorrect use of a query, our results did not show a significant decrease in the number of query errors by the sixth assessment.

Findings have important implications for professors teaching the WISC and strongly suggest that considerable time should be spent on this topic. This is consistent with the findings of Conner and Woodall (1983) who found that administration errors of graduate students significantly decreased, but only after receiving feedback about the type and number of errors committed. Given that the Similarities, Vocabulary, and Comprehension subtests represent core subtests and contribute to a child's General Ability Index (GAI), the importance of reducing errors on these subtests cannot be overstated. Errors could have significant implications for children referred for special education coding, especially with the recent emphasis of using the GAI in identifying learning disabilities (Bremner et al., 2011).

Compared to other studies, our results suggest that as a whole, only a small proportion of students were prone to making clerical/careless errors. Our results were much lower than previous studies (Belk et al., 2002; Loe et al., 2007; Slate et al., 1991). The importance of using correct normative tables and adding subtest and index scores was emphasized during the teaching of the WISC-IV. Perhaps a combination of making this a point of emphasis during classroom instruction and penalizing students for "human error" was enough to encourage students to double-check their work. In spite of this reasonably positive outcome, results suggest that graduate students are still prone to these types of careless errors, even after multiple administrations, and should be carefully monitored and checked for errors.

One of the most important outcomes of this study concerns the methodology of repeated practice to ensure improved student competence. Since the use of "practice makes perfect" mentality is intuitive, it continues to lack empirical support. In spite of the admonishment that students need at least five or six repetitions (Alfonso et al., 1998), the existing research fails to support this supposition. Previous studies using 5 administrations (Belk et al., 2002), 10 administrations (Slate et al., 1991; Slate & Jones, 1990), and even 15 administrations (Conner & Woodall) have not found improvements with repetition. The recent study by Loe et al. (2007) again found no significant improvements across three practice events. We sought to evaluate trends across six administrations in keeping with the recommendations by Alfonso et al., who suggested five or six administrations might be required. Our results suggest a mild trend for improvement across repeated trials. Of interest, in our study there was a statistically significant decrease in the number of errors from the first to third administration. Yet there appeared to be a "rebound effect" wherein student errors rebounded by the sixth administration. The explanation of this trend is speculative but may involve what we conceptualize as "examiner drift." It is possible that students were initially diligent to correct errors but became less attentive to learn from their mistakes with repeated trials. In our cohort, students were required to submit two reports by the end of the first term. This meant that there was a much heavier load of work during the second semester and we can only speculate that the combination of increased stress and work coupled with examiner drift contributed to these findings. In spite of using a form of negative punishment

(i.e., point deductions), students did not show the improvements that were expected with repeated administrations.

The problem of "examiner drift," however, is not unique to graduate students, as noted in studies with experienced professionals (Bradley et al., 1980; Sherrets et al., 1979). Thus, even seasoned examiners are prone to making similar errors to graduate students specifically on the verbal subtests of the WISC that require querying and subjective scoring. This finding underscores the importance of providing rigorous training methods to curtail examiner errors.

The findings from our study were consistent with past research and highlight the need for alternative training models. Outcomes from available research suggest that more passive approaches (expecting students to learn from their mistakes) have not helped reduce student errors. Although graduate students are believed to be highly motivated, this alone does not translate into decreasing errors with repeated administrations. Even punitive approaches (penalizing mistakes with reduced grades) did not yield significant reductions in administration and scoring errors. However, results from the current study suggest that the approach of penalizing students may have helped to reduce the number of clerical/careless errors.

To ensure professional competence in the administration of all psychological test measures, alternative delivery models should include more "hands on" approaches to training. This includes ensuring supervised instruction of graduate students provides a methodology that will establish professional competence. Consistent with the arguments of several studies headed by Slate (Slate & Jones, 1990; Slate et al., 1993) we recommend several important 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 supervisor/professor.

Other forms of feedback may also be helpful. For instance, Thompson and Hodgins (1994) developed an algorithm to enhance student scoring. This feedback tool was given to graduate students in addition to standard instructions about administration and scoring of the WAIS-R. Results were promising. To our knowledge, no other studies have been published developing other algorithms or procedures. This is likely due to the evolving computer scoring programs that have become available and the reliance on one's previous learning. Yet, consistent with the current body of research, such an approach is needed and could be potentially helpful. Future research could evaluate the effectiveness of the use of such an algorithm and the different methods recommended above.

Another possibility is enhancing the ease of scoring test items from the Wechsler scales. Changes to the WISC-IV and WAIS-IV included adding new subtests with more defined scoring outcomes (Picture Concepts) and removing other subtests outside the core battery. Not surprisingly, this change resulted in fewer errors. Yet the primary source of errors continues to be among Vocabulary, Similarities, and Comprehension subtests. The possibility of including items with more defined outcomes (several word responses) could improve scoring accuracy. Other intelligence

scales use this approach. However, this would need to be balanced with ensuring the validity of the constructs is accurately measured.

As with any study, there were weaknesses and limitations. The participants within this study came from several program streams (counseling psychology and school psychology) and, therefore, came with a diverse range of exposure to psychological testing. Since most students were in a master's program, it is likely they had a wide range of exposure and experience with psychological testing from their undergraduate programs. In addition, students did not have a prerequisite course to ensure their knowledge level was on a reasonably equal level at the start of the course. A second limitation was varying workloads during the course of the academic year. Some students had greater demands with practicum course work, and it is likely that time constraints were different. Whereas it is difficult to control a students' experience outside the classroom, the range of course demands could also have played a role in a student's stress levels and time availability for this course. Consistent with the Loe et al. study (2007), this study was restricted to the 10 core subtests and the student errors on other subtests could not be examined.

In summary, the findings from this study echo results from prior research that graduate students are prone to making a significant number of administration and scoring errors on the WISC-IV that does not appear to improve with practice. Alternative delivery models that address this concern need to be considered and empirically tested.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

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*(2), 119-125.

Belk, M. S., Lobello, 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.

Bradley, F. O., Hanna, G. S., & Lucas, B. A. (1980). The reliability of scoring of the WISC-R. *Journal of Counseling and Clinical Psychology, 48*(4), 530-531.

Bremner, D., McTaggart, B., Saklofske, D. H., & Janzen, T. L. (2011). WISC-IV GAI and CPI in psychoeducational assessment. *Canadian Journal of School Psychology, 26*, 209-219.

Conner, R., & Woodall, F. E. (1983). The effects of experience and structured feedback on WISC-R error rates made by student-examiners. *Psychology in the Schools, 20*, 376-379.

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, 23*(3), 237-247.

Miller, C. K., & Chansky, N. M. (1972). Psychologists' scoring of WISC protocols. *Psychology in the Schools, 9*, 144-152.

Ryan, J. J., Prifitera, A., & Powers, L. (1983). Scoring reliability of WISC protocols. *Psychology in the Schools, 9*, 144-152.

Sherrets, S., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. *Psychology in the Schools, 16*, 495-496.

Slate. J. R., & Chick, D. (1989). WISC-R examiner errors: Cause for concern. *Psychology in the Schools, 26*, 78-84.

Slate, J. R., & Jones, C. H. (1990). Student error in administering the WISC-R: Identifying problem areas. *Measurement and Evaluation in Counseling and Development, 23*, 137-140.

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 administration. *Professional Psychology: Research and Practice, 22*, 375-379.

Slate, J. R., Jones, C. H., & Murray, R. A. (1993). Evidence that practitioners err in administration and scoring of the WAIS-R. *Measurement and Evaluation in Counseling and Development, 25*, 156-161.

Thompson, A., & Hodgins, C. (1994). Evaluation of a checking procedure for reducing clerical and computational errors on the WAIS-R. *Canadian Journal of Behavioral Science, 26*, 492-504.

Wechsler, D. (2003). *Wechsler intelligence scale for children* (4th ed.). San Antonio, TX: The Psychological Corporation.

Bios

Martin Mrazik, PhD, is associate professor and director of the Assessment Center in the Department of Educational Psychology at the University of Alberta. He is a clinical neuropsychologist with special interest in children with neurodevelopmental disabilities and acquired brain injury.

Troy M. Janzen, PhD, RPsych, is an adjunct assistant professor and clinical supervisor for the Professional Studies in Education graduate programs at the University of Alberta. His research interests include cognitive assessment, graduate training in assessment, reading and reading disability, and studying these factors among First Nations populations. He currently serves as the editor of the joint newsletter for the Psychologists in Education and Canadian Association of School Psychologists.

Stefan C. Dombrowski, PhD, is professor and director of the School Psychology program at Rider University. He was the 2011 recipient of Rider University's Dominick A. Iorio distinguished research award. He is a licensed psychologist and certified school psychologist who has written three books and more than three-dozen articles.

Sean Barford is a fourth-year PhD student in counseling psychology at the University of Alberta. He is currently specializing in psychoeducational, personality, cognitive, and parental assessment while completing his clinical internship in the Clinical Services department at the University of Alberta.

Lindsey Krawchuk is a PhD student in educational psychology at the University of Alberta. Her research interests include student motivation, school psychology, and addiction and mental health.