Archives of Clinical Neuropsychology Advance Access published October 13, 2014

OXFORD
UNIVERSITY PRESS

Archives of CLINICAL NEUROPSYCHOLOGY

Comparing Canadian and American Normative Scores on the Wechsler Adult Intelligence Scale-Fourth Edition

Allyson G. Harrison[1,*], Irene T. Armstrong[1], Laura E. Harrison[1], Rael T. Lange[2,3], Grant L. Iverson[4,5]

[1]*Regional Assessment & Resource Centre, Queens University, Kingston, ON, Canada*
[2]*Defense and Veterans Brain Injury Center, Walter Reed National Military Medical Center, Bethesda, MD, USA*
[3]*University of British Columbia, Vancouver, Canada*
[4]*Department of Physical Medicine and Rehabilitation, Harvard Medical School, & Red Sox Foundation and Massachusetts General Hospital Home Base Program, Boston, MA, USA*
[5]*Copeman Healthcare Centre, Vancouver, BC, Canada*

*Corresponding author at: Regional Assessment & Resource Centre, Queen's University, Mackintosh-Corry Hall, B10068, University Avenue, Kingston, ON, Canada K7L 3N6. Tel.: +1-613-533-6311
E-mail address: harrisna@queensu.ca (A. G. Harrison).

Accepted 8 September 2014

Abstract

Psychologists practicing in Canada must decide which set of normative data to use for the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). The purpose of this study was to compare the interpretive effects of applying American versus Canadian normative systems in a sample of 432 Canadian postsecondary-level students who were administered the WAIS-IV as part of an evaluation for a learning disability, attention-deficit hyperactivity disorder, or other mental health problems. Employing the Canadian normative system yielded IQ, Index, and subtest scores that were systematically lower than those obtained using the American norms. Furthermore, the percentage agreement in normative classifications, defined as American and Canadian index scores within five points or within the same classification range, was between 49% and 76%. Substantial differences are present between the American and Canadian WAIS-IV norms. Clinicians should consider carefully the implications regarding which normative system is most appropriate for specific types of evaluations.

Keywords: Wechsler adult intelligence scale-IV; Normative systems; Canadian norms; Assessment

Introduction

Canadian psychologists face the perplexing issue of whether to use the Canadian normative data or the American normative data for the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008a, 2008b). The American normative data for this test are based on a much larger sample, it is co-normed with other test batteries (such as the Wechsler Memory Scale-Fourth Edition and the Wechsler Individual Achievement Test-Second Edition), it includes both age and demographically adjusted normative scores, and numerous studies (e.g., Chaudhry & Ready, 2012; Heyanka, Holster, & Golden, 2013; Nelson, Canivez & Watkins, 2013) and book chapters (e.g., Brooks et al., 2013; Cullum & Lacritz, 2009; 2012; Cullum & Larabee, 2010; Holdnack, Schoenberg, Lange, & Iverson, 2013) are published that use these normative data. Furthermore, the majority of neuropsychological and achievement tests offer only American norms. Hence, clinicians in Canada have to choose whether to combine scores from different normative samples when assessing a client, or convert all the raw scores in an assessment using a common metric (i.e., American normative data) and thus also avail themselves of an extensive body of research regarding known score patterns for specific disorders.

This issue has been a challenge for >25 years since Canadian psychologists have questioned the appropriateness of American normative data for use in Canada (e.g., Beal, 1988). This concern prompted the Psychological Corporation (now Pearson) to evaluate whether data from the two countries were equivalent, ultimately leading the test publisher to develop separate Canadian normative data (e.g., Wechsler, 1996, 2001, 2003, 2004). Initially, separate normative data were obtained in the updated version of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991, 1996). This was followed by Canadian norms being developed for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 2001) and the WAIS-IV (Wechsler, 2008b).

doi:10.1093/arclin/acu048

On the WISC-III and the WISC-IV, researchers have reported that Canadian children obtain higher raw scores than American children (Beal, Dumont, Cruse & Branche, 1996; Wechsler, 2004). Similarly, on the WAIS-III and the WAIS-IV, researchers have reported that Canadian adults obtain higher raw scores than American adults, although these differences are less apparent in adults over the age of 65 (Wechsler, 2008b). The reasons for these differences are unclear. The developers of the Wechsler scales hypothesized that these differences emerged in part due to variations in population composition between the two countries with respect to race, ethnicity, and educational attainment (WAIS-IV Canadian technical manual, Wechsler, 2008b). This hypothesis was supported by data from the WISC-IV standardization illustrating that all significant differences in obtained composite scores disappeared (except for processing speed) when samples from the two countries were matched on key demographic characteristics (Wechsler, 2004). Although not investigated empirically, it was suggested that various social, economic, and educational differences between Canada and the United States may account for a significant portion of the discrepancy between the Full Scale IQ (FSIQ) scores of Canadian and American children (Wechsler, 2004, p. 43). The manual notes that Canadian children are typically required to begin school 1 year earlier than do most American children, and that a higher percentage of Canadian versus American parents/guardians have some college-level education.

There are also substantial statistical differences in how the American versus Canadian norms were developed that could contribute to the differences between normative scores. For example, the American WAIS-IV standardization sample comprises 2,200 individuals between ages 16–90, stratified into 13 age groups. In contrast, the Canadian norms were based on a sample of 688 individuals. Norms for both versions were generated using a complex statistical technique called continuous norming (Gorsuch, 2003; Wechsler, 2001), a norming method developed to mitigate the effects of relatively small sample sizes across age groups. Continuous norming has been recommended to correct for irregularities in either the distributions of scores within groupings of the norming variable or to calculate trends in the means and standard deviations across groupings when group sample sizes are 200 or smaller (Angoff & Robertson, 1987).

Two research groups have previously examined the clinical implications of using American versus Canadian normative data for psychologists practicing in Canada. Beal, Dumont, Cruse, and Branche (1996) examined the performance of children with learning disabilities on the WISC-III. They found scores to be systematically lower when employing the Canadian norms for Verbal IQ, FSIQ, Verbal Comprehension (VCI), Perceptual Organization (POI), and Freedom from Distractibility Indexes. Lowered scores resulted in changes to FSIQ classifications in ~32% of the sample, although they stressed that use of the confidence interval mitigated the effects of these scoring changes. Iverson, Lange, and Viljoen (2006) analyzed the effect that using different normative systems had on the obtained WAIS-III scores of 100 forensic psychiatry and neuropsychiatry inpatients. Use of the Canadian normative data consistently lowered the obtained scores for the same individual relative to scoring the protocol using American norms, frequently causing the Index scores to change classification status (e.g., lowering obtained IQ from the Low Average to the Borderline range). Scores that differed most significantly were FSIQ and Performance IQ, while the Processing Speed Index (PSI) and VCI were most similar using both sets of normative data. The authors noted that clinicians evaluating patients with acquired cognitive deficits might conclude that a person has a greater degree of cognitive impairment when using the Canadian norms compared with the American norms. Moreover, Iverson and colleagues cautioned clinicians about the implications of mixing Canadian-derived IQ scores with American-derived achievement tests or vice versa, and encouraged practitioners to have good a priori reasons for choosing one set of normative data over the other.

Given that previous research with the WAIS-III and the WISC-III revealed systematic differences in obtained Index and Scaled Scores depending on which set of normative data was used, the present study sought to examine the effect of using different normative data when calculating scores on the WAIS-IV. Such information is particularly important because norming differences may influence score interpretation and classification of clients (Iverson, Lange, & Viljoen, 2006; Ryan & Schnakenberg-Ott, 2003), especially when IQ cut scores are employed in specific diagnostic classification schemes (e.g., IQ-Achievement discrepancy and the diagnosis of Learning Disability [Cotton, Crewther, & Crewther, 2005; Willis & Dumont, 2006]; identification of Intellectual Disability [Harrison & Holmes, 2014]). It was hypothesized that there would be statistically significant differences on all IQ, Index, and subtest scaled scores, with Canadian normative scores being lower than American normative scores. It was also hypothesized that a substantial minority of people would have clinically meaningful differences in normative scores when comparing results from the two systems.

Method

Participants

Participants in this ethics-approved study were 432 students who completed the WAIS-IV as part of a clinical evaluation provided by a regional assessment center serving students in the southern half of Ontario. All participants had been referred for a psycho-educational or neuropsychological evaluation of learning and/or attention problems, and had given consent to have their data used for research purposes. Students referred to this assessment service had either been accepted into or were currently studying at a postsecondary institution somewhere in the province of Ontario. The primary purpose of the referrals was to determine if they qualified for a disability diagnosis that would allow them to obtain academic accommodations and supports at their home institution. Although many had come to postsecondary studies directly (i.e., after having graduated high school), a sizeable minority were returning as mature students, often after completing an academic upgrading or a high school equivalency program. The sample also included a small proportion of graduate students referred due to recently reported academic difficulties.

Their ages ranged from 16 to 57 years (mean = 22.6 years, *SD* = 6.5), and the majority of the sample was female (56%). Although the breakdown of race and ethnicity was not available, it is known from referral characteristics that the majority of students were Caucasian. Many of the referred students received a diagnosis of a specific learning disability (38.2%) or attention-deficit hyperactivity disorder (ADHD; 23.9%), including 6.6% diagnosed with both a learning disability and ADHD. However, many others were found to have learning or attention problems due to other causes such as generalized anxiety, depression, borderline intellectual functioning, obsessive–compulsive personality disorder, perfectionism, or weak academic background. In fact, 14.3% of the sample obtained a diagnosis other than LD or ADHD, and 23.3% received no diagnosis. The sample, therefore, was rather heterogeneous both in terms of final diagnosis and level of intellectual ability.

Procedure

To evaluate the difference in obtained scores when applying either the Canadian or American WAIS-IV normative systems, all protocols were scored using the commercially available computer scoring program (WAIS-IV Scoring Assistant software), which allows for the generation of standard scores using either Canadian or American norms. Protocols with score differences >1 *SD* were rescored to ensure accurate calculation of index and scaled scores.

Results

Descriptive statistics, correlation coefficients, mean comparisons and effect sizes for the Composite, Index, and subtest scores using the American and Canadian WAIS-IV normative systems are presented in Table 1. The Pearson correlations between the two normative systems for all Composite (i.e., FSIQ and General Ability Index [GAI]) and Index scores was $r = .99$. For the subtest scores, correlation coefficients ranged from $r = .96$ to .99. Intraclass correlation coefficients for the Composite, Index, and subtest scores were similarly high (range: $r = .95$–.99). Using paired sample *t*-tests, significantly lower scores ($p < .001$) were found on all Composite, Index, and subtest scores when using the Canadian norms compared to the American norms. Medium effect sizes were found when comparing FSIQ, GAI, Working Memory Index (WMI), Digit Span, and Letter Number Sequencing (range: $d = .43$–.53). Effect sizes for the remaining Index and subtests scores range from small ($d = .21$) to small-medium ($d = .38$).

Difference scores between Canadian- and American-generated Composite, Index, and subtest scores were calculated by subtracting corresponding values between the normative systems. The mean difference scores and percent agreement between the two normative systems are presented in Table 2 (positive difference scores reflect higher American scores). The vast majority of the sample had higher Composite scores (FSIQ = 99.1%, GAI = 98.7%), Index scores (VCI = 98.1%, POI = 99.3%, WMI = 99.5%, PSI = 92.3%), and subtests scores (range: 63% to 95%) when using the American norms. The percent agreement between the two normative systems was calculated using three criteria: first, percent within one third of a *SD* (i.e., ± 5 points for Composite and Index scores; 1 point for subtest scores); secondly, percent within the same ability classification level (ranging from extremely low to very superior); thirdly, percent within one third of a *SD* or within one ability classification. The highest rate of agreement (within one third of a *SD* or the same classification range) across the Composite and Index scores was found for POI (76.3%) and PSI (72.6%). Relatively low rates of agreement were found for the remaining Composite and Index scores (range = 49.0%–66.1%). All of the subtest scores, with the exception of Digit Span and Letter Number Sequencing, had higher rates of agreement. The highest rates of agreement were found for the Information (99.5%), Coding (99.1%), Arithmetic (97.7%), and Matrix Reasoning (94.9%) subtests.

In order to explore the influence of overall intellectual ability on score differences between normative systems, the difference scores and percent agreement for the Composite, Index, and subtests scores, stratified by American FSIQ, are presented in Tables 3 and 4. A clear trend was evident whereby Composite and Index score differences were consistently larger as overall FSIQ score decreased. Thus, while almost all subjects obtained a lower score when the Canadian normative system was applied,

Table 1. Descriptive statistics, correlations, mean comparisons, and effect sizes

Score	American norms		Canadian norms		p	d	r	ICC
	M	SD	M	SD				
Composite Scores								
FSIQ	98.4	12.9	91.2	14.6	*	0.52	0.99	0.99
General Ability Index ($n = 398/423$)	102.4	14.7	95.8	15.8	*	0.43	0.99	0.99
Index Scores								
Verbal Comprehension	101.5	15.0	95.6	16.1	*	0.38	0.99	0.99
Perceptual Reasoning	102.0	14.0	96.8	15.8	*	0.35	0.99	0.99
Working Memory	92.4	12.7	85.4	13.8	*	0.53	0.99	0.99
Processing Speed	96.4	13.0	91.5	14.5	*	0.36	0.99	0.98
Subtest Scores								
Verbal Subtests								
Vocabulary	10.5	3.2	9.3	3.4	*	0.36	0.99	0.98
Similarities	10.4	2.8	9.3	3.1	*	0.37	0.98	0.98
Information	9.9	3.2	9.2	3.4	*	0.21	0.99	0.99
Arithmetic	8.6	2.8	7.7	2.8	*	0.32	0.99	0.99
Digit Span	8.6	2.6	7.4	2.7	*	0.45	0.98	0.97
Letter Number Sequencing	9.0	2.3	7.8	2.5	*	0.50	0.96	0.95
Performance Subtests								
Block Design	10.1	3.0	9.2	3.2	*	0.29	0.99	0.99
Matrix Reasoning	10.4	2.9	9.7	3.2	*	0.23	0.99	0.98
Visual Puzzles	10.7	2.9	9.6	3.0	*	0.37	0.95	0.95
Symbol Search	9.5	2.8	8.7	3.0	*	0.28	0.99	0.98
Coding	9.2	2.6	8.4	2.6	*	0.31	0.97	0.97

Notes: $N = 428–435$ for all variables unless shown in parentheses beside variable name. Within parentheses, American n/Canadian n.
*$p < .001$. http://www.uccs.edu/~lbecker/ for calculation of d (and r effect size—see Jan 2513).The variable r is correlation coefficient.

Table 2. Mean differences and percentage agreement among systems: Composite, Index, and subtest scores

	Mean difference (SD)	% within ±5 points	% within same classification	% within ±5 points or same classification
Composite Scores				
FSIQ	7.2 (2.3)	23.2	42.5	49.7
General Ability Index	6.2 (2.2)	39.2	49.1	58.8
Index Scores				
Verbal Comprehension	5.9 (2.4)	43.2	54.1	66.1
Perceptual Reasoning	5.2 (2.6)	60.6	65.2	76.3
Working Memory	7.0 (2.3)	24.1	45.5	49.0
Processing Speed	4.8 (2.6)	54.1	60.3	72.6
	Mean difference (SD)	% within ±1 point	% within same classification	% within ±1 point or same classification
Verbal Subtests				
Vocabulary	1.2 (0.6)	72.6	58.4	76.0
Similarities	1.2 (0.7)	72.6	63.8	81.7
Information	0.6 (0.5)	99.3	80.5	99.5
Arithmetic	0.9 (0.6)	97.7	70.6	97.7
Digit Span	1.2 (0.6)	70.4	51.7	74.1
Letter Number Sequencing	1.1 (0.7)	70.2	50.0	71.1
Performance Subtests				
Block Design	0.9 (0.5)	91.9	64.5	91.9
Matrix Reasoning	0.7 (0.6)	94.7	78.0	94.9
Visual Puzzles	1.1 (0.9)	79.0	62.5	80.4
Symbol Search	0.8 (0.5)	94.2	79.1	94.7
Coding	0.8 (0.7)	99.1	75.4	99.1

Notes: Difference is American Norms minus Canadian Norms; positive values indicate greater American norms.

significantly larger differences were found for individuals the farther their FSIQ score was below the mean. Furthermore, as shown in Table 4, the percentage of individuals whose scores would be interpreted as clinically similar (i.e., within the same classification level) also decreased as FSIQ level declined. These findings were true for all Composite and Index scores.

Table 3. Mean and *SD* of differences between American–Canadian norms for Composite, Index, and Subtest scores by American FSIQ level

	American FSIQ Level						
	<70 ($n = 1$) *M*	70–79 ($n = 32$) *M (SD)*	80–89 ($n = 78$) *M (SD)*	90–109 ($n = 241$) *M (SD)*	110–119 ($n = 53$) *M (SD)*	120–130 ($n = 20$) *M (SD)*	>130 ($n = 7$) *M (SD)*
Composite Scores							
FSIQ	11.0	9.3 (1.2)	8.9 (1.5)	7.2 (2.1)	4.9 (1.4)	4.8 (2.1)	5.0 (1.0)
GAI	10.0	8.3 (1.5)	7.9 (1.7)	5.7 (2.1)	5.3 (1.7)	4.9 (2.8)	4.3 (0.8)
Index Scores							
Verbal Comprehension	7.0	7.6 (2.3)	7.0 (1.8)	5.5 (2.2)	5.3 (2.4)	4.7 (4.1)	4.3 (1.3)
Perceptual Reasoning	11.0	8.8 (1.6)	7.1 (2.2)	4.6 (2.3)	3.8 (1.6)	4.0 (1.5)	3.6 (1.3)
Working Memory	9.0	7.9 (2.1)	7.8 (2.1)	6.9 (2.5)	6.5 (1.7)	5.4 (1.6)	4.9 (0.9)
Processing Speed	7.0	6.7 (2.1)	5.6 (2.1)	4.7 (2.4)	3.8 (2.7)	3.1 (3.6)	3.3 (2.0)
Verbal Subtests							
Vocabulary	2.0	1.6 (0.5)	1.4 (0.7)	1.1 (0.5)	1.2 (0.6)	1.0 (0.6)	1.0 (0.6)
Similarities	1.0	1.3 (0.7)	1.4 (0.5)	1.1 (0.7)	1.0 (0.6)	0.9 (0.6)	0.4 (0.5)
Information	0.0	0.9 (0.3)	0.8 (0.6)	0.6 (0.5)	0.6 (0.5)	0.4 (0.6)	0.4 (0.5)
Arithmetic	1.0	1.0 (0.5)	0.9 (0.6)	0.9 (0.5)	0.9 (0.3)	0.9 (0.3)	1.0 (0.0)
Digit Span	2.0	1.3 (0.5)	1.3 (0.7)	1.2 (0.6)	1.2 (0.5)	1.1 (0.4)	0.9 (0.4)
Letter Number Sequencing	1.0	1.4 (0.7)	1.3 (0.6)	1.0 (0.7)	0.8 (0.9)	1.0 (0.8)	
Performance Subtests							
Block Design	2.0	1.2 (0.6)	1.0 (0.6)	0.8 (0.5)	0.8 (0.4)	1.0 (0.2)	0.7 (0.5)
Matrix Reasoning	1.0	1.1 (0.6)	1.0 (0.5)	0.6 (0.6)	0.4 (0.5)	0.5 (0.5)	0.7 (0.5)
Visual Puzzles	2.0	1.6 (0.5)	1.3 (0.6)	1.0 (0.7)	1.3 (0.4)	0.2 (3.1)	0.6 (0.5)
Symbol Search	1.0	1.0 (0.5)	0.9 (0.5)	0.9 (0.5)	0.7 (0.5)	0.7 (0.7)	0.6 (0.5)
Coding	1.0	0.9 (0.3)	0.8 (0.5)	0.7 (0.5)	0.8 (1.5)	0.7 (0.5)	0.9 (0.4)

Notes: There is only one client with FSIQ <70 using American norms, thus, no *SD*. All scores significant at $p \leq .001$, except as follows: $p \leq .01$, $p \leq .05$, $p =$ n.s.

Table 4. Percentages of subjects with clinically similar scores between normative systems as a function of American FSIQ level

	American FSIQ level						
	< 70 ($n = 1$)	70–79 ($n = 32$)	80–89 ($n = 78$)	90–109 ($n = 241$)	110–119 ($n = 53$)	120–130 ($n = 20$)	>130 ($n = 7$)
Composite Scores							
FSIQ	100.0	12.5	15.6	59.3	66.0	70.0	71.4
GAI	100.0	31.0	20.3	70.8	61.2	66.7	100.0
Index Scores							
Verbal Comprehension	100.0	28.1	43.6	74.7	75.5	75.0	100.0
Perceptual Reasoning	100.0	12.5	51.3	86.7	96.2	90.0	100.0
Working Memory	100.0	25.0	29.5	49.0	66.0	95.0	100.0
Processing Speed	100.0	40.6	61.0	75.1	86.8	90.0	100.0
Verbal Subtests							
Vocabulary	100.0	37.5	57.7	84.6	83.0	80.0	85.7
Similarities	100.0	56.3	71.8	86.0	86.8	90.0	100.0
Information	100.0	100.0	97.4	100.0	100.0	100.0	100.0
Arithmetic	100.0	90.6	97.4	97.9	100.0	100.0	100.0
Digit Span	100.0	65.6	64.5	74.3	81.1	90.0	100.0
Letter Number Sequencing	100.0	53.8	62.1	79.3	70.0	75.0	
Performance Subtests							
Block Design	100.0	71.9	83.3	95.0	100.0	100.0	100.0
Matrix Reasoning	100.0	84.4	87.2	96.7	100.0	100.0	100.0
Visual Puzzles	100.0	43.8	69.2	85.8	88.7	90.0	100.0
Symbol Search	100.0	87.5	92.1	95.9	96.2	95.0	100.0
Coding	100.0	100.0	98.7	99.2	98.1	100.0	100.0

To explore the influence of age on score differences between normative systems, the difference scores and percent agreement for the Composite, Index, and subtests scores, stratified by age cohorts, are presented in Tables 5 and 6. For all Composite and Index scores (except the PSI), 100% of the younger subjects obtained higher scores using the American norms (Table 5). For the VCI, it is only in the 35- to 44-year-old age group that a few subjects obtained higher scores using the Canadian norms,

Table 5. Mean and *SD* of differences between American–Canadian norms for Composite, Index, and Subtest scores by age

	Age (years)							
	16–17 ($n = 13$) *M* (*SD*)	18–19 ($n = 157$) *M* (*SD*)	20–24 ($n = 119$) *M* (*SD*)	25–29 ($n = 44$) *M* (*SD*)	30–34 ($n = 14$) *M* (*SD*)	35–44 ($n = 16$) *M* (*SD*)	45–54 ($n = 6$) *M* (*SD*)	55–64 ($n = 2$) *M* (*SD*)
Composite Scores								
FSIQ	7.6 (1.7)	7.7 (1.9)	7.0 (2.2)	7.6 (2.5)	6.8 (2.2)	4.4 (2.7)	2.3 (3.1)	2.5 (2.1)
General Ability Index	6.2 (1.4)	6.3 (2.0)	6.1 (2.0)	6.5 (2.2)	5.8 (2.4)	4.2 (3.1)	2.0 (2.3)	2.5 (0.7)
Index Scores								
Verbal Comprehension	5.8 (2.3)	6.5 (2.1)	5.6 (2.0)	5.4 (2.5)	5.9 (2.7)	2.9 (3.6)	2.2 (2.3)	0.0 (0.0)
Perceptual Reasoning	5.5 (2.4)	5.0 (2.5)	5.2 (2.5)	6.8 (3.0)	5.4 (2.3)	4.6 (2.8)	2.5 (2.4)	3.0 (0.0)
Working Memory	5.4 (2.1)	7.1 (2.2)	7.3 (1.9)	7.7 (2.2)	5.7 (1.4)	4.2 (2.4)	2.8 (1.2)	4.5 (3.5)
Processing Speed	5.4 (2.3)	5.4 (2.7)	4.8 (2.3)	4.9 (2.3)	4.8 (1.6)	3.4 (3.2)	1.0 (2.5)	0.5 (2.1)
Verbal Subtests								
Vocabulary	1.2 (0.6)	1.4 (0.6)	1.2 (0.5)	1.0 (0.4)	1.1 (0.4)	0.7 (0.8)	0.5 (0.5)	0.0 (0.0)
Similarities	1.2 (0.8)	1.3 (0.5)	1.1 (0.5)	0.9 (0.8)	1.1 (0.9)	0.8 (1.5)	0.7 (0.5)	0.0 (0.0)
Information	0.6 (0.7)	0.7 (0.6)	0.6 (0.5)	0.7 (0.5)	0.7 (0.5)	0.5 (0.6)	0.0 (0.0)	0.0 (0.0)
Arithmetic	1.1 (1.0)	0.8 (0.5)	0.9 (0.4)	1.1 (0.5)	0.9 (0.3)	0.5 (0.5)	0.8 (0.4)	1.0 (0.0)
Digit Span	0.8 (0.8)	1.3 (0.6)	1.3 (0.5)	1.3 (0.7)	0.9 (0.3)	0.7 (0.6)	0.0 (0.0)	0.5 (0.7)
Letter Number Seq	2.0 (1.0)	1.2 (0.7)	0.8 (0.7)	0.8 (0.7)	1.0 (0.0)			
Performance Subtests								
Block Design	0.8 (0.4)	0.8 (0.5)	1.9 (0.5)	1.1 (0.6)	0.6 (0.6)	0.8 (0.5)	0.7 (0.5)	1.0 (0.0)
Matrix Reasoning	0.8 (0.6)	0.7 (0.6)	0.7 (0.6)	1.0 (0.5)	1.1 (0.4)	0.5 (0.6)	0.2 (0.4)	0.5 (0.7)
Visual Puzzles	1.1 (0.5)	1.1 (0.5)	1.1 (0.5)	1.4 (0.6)	0.4 (1.9)	1.2 (0.5)	0.2 (0.4)	0.0 (0.0)
Symbol Search	0.7 (0.8)	1.0 (0.5)	0.8 (0.4)	0.8 (0.5)	0.7 (0.5)	0.6 (0.5)	−0.3 (0.8)	0.0 (0.0)
Coding	0.9 (0.3)	0.9 (0.9)	0.8 (0.5)	0.8 (0.4)	0.9 (0.3)	0.4 (0.5)	0.5 (0.5)	0.0 (0.0)

Notes: Positive scores indicate American Norms greater than Canadian Norms. All scores significant at $p \leq 0.001$, except as follows: $p \leq .01$, $p \leq .05$, $p =$ n.s.

Table 6. Percentages of subjects with clinically similar scores between normative systems by age

	Age (years)							
	16–17 ($n = 13$) %	18–19 ($n = 157$) %	20–24 ($n = 120$) %	25–29 ($n = 44$) %	30–34 ($n = 14$) %	35–44 ($n = 17$) %	45–54 ($n = 6$) %	55–64 ($n = 2$) %
Composite Scores								
FSIQ	30.8	41.4	55.5	54.5	64.3	75.0	83.3	100.0
GAI	45.5	58.7	61.4	59.0	69.2	62.5	100.0	100.0
Index Scores								
Verbal Comp.	76.9	59.2	70.0	77.3	64.3	76.5	100.0	100.0
Perceptual Reasoning	61.5	76.4	77.5	63.6	78.6	82.4	83.3	100.0
Working Memory	61.5	47.1	44.5	45.5	57.1	76.5	100.0	100.0
Processing Speed	53.8	65.6	75.6	79.5	85.7	81.3	100.0	100.0
Verbal Subtests								
Vocabulary	69.2	67.5	80.0	90.9	85.7	88.2	100.0	100.0
Similarities	76.9	88.0	80.2	75.0	57.1	94.1	100.0	100.0
Information	100.0	98.7	100.0	100.0	100.0	100.0	100.0	100.0
Arithmetic	91.7	98.7	99.2	86.4	100.0	100.0	100.0	100.0
Digit Span	100.0	68.2	75.8	72.7	100.0	94.1	100.0	100.0
Letter Number Seq.	66.7	63.8	83.3	69.2	100.0	100.0	100.0	
Performance Subtests								
Block Design	100.0	94.3	89.3	79.5	92.9	94.1	100.0	100.0
Matrix Reasoning	100.0	96.2	93.4	88.6	85.7	94.1	100.0	100.0
Visual Puzzles	83.3	82.9	83.9	63.6	92.9	88.2	100.0	100.0
Symbol Search	91.7	89.2	100.0	97.7	100.0	100.0	100.0	100.0
Coding	100.0	98.1	99.1	100.0	100.0	100.0	100.0	100.0

Notes: For composite scores and indexes: % within ±5 points or same classification; for subtests: % within ±1 point or same classification.

and for FSIQ it is only in the 45- to 54-year-old group that a slight advantage (1 IQ point) was obtained for one subject when applying the Canadian norms. Not one subject returned a higher score using Canadian norms on the POI or WMI nor was there a single individual who returned a higher score on the GAI when Canadian norms were employed.

Some large differences in scores were found for a few individuals in the sample. For example, there were times when the difference in computed Composite or Index score exceeded 1 *SD*, with the Canadian score being 18 points lower than the American score for VCI (data not shown but available on request). The largest difference for POI was 13 points, WMI was 14 points, PSI 13 points, FSIQ 13 points, and GAI 13 points.

Discussion

This study investigated the effect of using Canadian versus American normative data when calculating scores on the WAIS-IV. Although the scores produced by the two systems were very highly correlated, systematically lower scores were obtained when using Canadian norms. Indeed, the results of our investigation confirm that substantially different scores were produced depending on the norms employed, and that the effect size of this difference was moderate. Overall, significantly lower scores were obtained on all Composite, Index, and subtest scores when employing Canadian as opposed to American norms. The largest differences were found in FSIQ and WMI, and the smallest score difference was found in the PSI score. Even so, clinically different conclusions (e.g., changing IQ or index score classifications from Below Average to Borderline, or the like) might be reached for between 25% and 51% of these students. Half of the individuals had clinically meaningful differences in their FSIQ scores across the two normative systems (i.e., greater than ± 5 points and a different classification).

Contrary to the findings of Iverson, Lange, and Viljoen (2006), the rates of agreement between these two normative systems when examining data at the subtest level were more variable, with between 1%–29% of students in the present study receiving different scores depending on the system employed, whereas Iverson et al. (2006) found that 0%–9% received different scores. Assuming that scores within ± 1 point or within the same classification qualify as agreement between systems, the subtests that differed most were Letter–Number Sequencing (71% agreement), Digit Span (74% agreement), and Vocabulary (76% agreement). Given that the trend towards individuals receiving lower scores on all Composite, Index, and subtests when Canadian norms were applied, it was not surprising to find that individuals had more scores in the impaired range when this normative system was used. Indeed, while only 10% of the sample had 5 or more low scores (Composite or Index score <80 or a scaled score of 6 or less) using the American norms, 24% had that many low score when their protocols were scored using the Canadian norms. Hence, clinicians would conclude greater impairment in intellectual abilities if Canadian norms were used. The differences in the normative scores derived from the two systems might be due to a real difference between the two normative samples, statistical differences in the methods by which each set of normative data were calculated, or both.

This study also identified that larger discrepancies between American- and Canadian-derived scores were found in those individuals whose IQ's were <100, and that the size of the difference in calculated Composite and Index scores increased significantly as FSIQ decreased. That is, those with higher FSIQ's did not experience the same magnitude of difference in their scores than did those participants whose FSIQ's were <100. In addition, the rate of general agreement in classification between the two normative systems was also significantly lower in those whose FSIQ scores were <100. In other words, clinicians would be more likely to draw different clinical conclusions about the meaning of IQ, index or subtest scores for individuals with American FSIQs of 100 or less. This finding has major implications for the number of individuals who would be considered to potentially meet the IQ criteria for an Intellectual Disability (i.e., an FSIQ score at or below the 2nd percentile). Indeed, using the American normative system, 3% of the individuals in our sample obtained an FSIQ score of 75 or less (using the cut score of FSIQ = 70 ± 5 to account for measurement error). In contrast, 15% of the individuals in our sample would obtain an FSIQ score in this range when the Canadian normative system was applied. While diagnosis of Intellectual Disabilities is not supposed to rely on an IQ score alone (see Harrison & Holmes, 2014), this score is "routinely used to classify mental disability" (McDermott, Watkins, & Rhoad, 2014, p. 207). Further, it is the case that in Ontario, Canada, students are routinely identified as having a "Mild Intellectual Disability" simply if their FSIQ score falls <80 or 85 (Harrison & Holmes, 2014). While only 17.5% of the students in the current sample returned an FSIQ of 85 or less using American norms, 40% had a score this low when Canadian norms were employed. Such scoring differences would have broad implications for the number of postsecondary students in Canada who would qualify as having a disability. Indeed, given that individuals identified as having an Intellectual Disability are eligible for many government-sponsored programs, services, and tax exemptions, the difference in classification rates has personal, social, and political ramifications. Worth noting, too, is the fact that all of the students in the present sample had been accepted into or were currently participating in a postsecondary program of study. None were in modified programs designed for persons with Intellectual Disabilities. While it is possible that a student with a bona fide Intellectual Disability might have chosen to return to postsecondary studies as mature students, it seems difficult to believe that up to 40% of these postsecondary students would have met the entrance requirements at their respective institutions and yet were significantly impaired intellectually. One must therefore question whether the Canadian normative data may contribute to overdiagnosis of Intellectual Disability; more research is needed to examine this issue.

When Canadian norms were applied, only one individual in the sample obtained an FSIQ score that was higher (by one point) than the corresponding American normed score. Furthermore, on all other Index scores, younger students in particular (those under age 35 or 45) always obtained lower scores when Canadian norms were applied. Hence, it appears that the Canadian norms systematically and consistently return lower scores for all individuals, especially young adults and those whose abilities are below average (i.e., <100). On average, the entire sample was about half a *SD* lower than the American norms. Large difference scores occurred exclusively in the students under age 29, most often in the 16- to 19-year-old age group with some scores changing more than a *SD*, although, for the younger age group, smaller cell sizes likely represent unstable findings that require further research.

We can only speculate as to the reasons for these observed differences. As Kahneman (2011) notes, large samples yield data that are more precise than do smaller samples, and smaller samples yield extreme results more often than do larger samples. The Canadian sample size was relatively small and those in the 16- to 17-year-old group appear to have an overrepresentation of individuals from all racial/cultural backgrounds with higher than average education. Similarly, the 18- to 24-year-old sample had an overrepresentation of persons of Asian descent with higher levels of education (Wechsler, 2008b, p. 43). Individuals who identified as being of First Nations decent appear to have been recruited primarily from large urban centers rather than from more remote reserves, which may also have led to sampling bias and the total number of First Nations persons included in the 16- to 19-year-old age range was significantly lower than the proportion suggested by the Canadian census. It therefore seems possible that some sampling bias was present that could explain why the raw scores obtained during the norming were higher in these age groups. Given that the transformation of the Canadian raw data was based on the American performance curves, it is possible that this non-representative and more highly educated Canadian sample were penalized when the raw data was fitted to the normal curve. Similarly, if fewer lower IQ individuals were actually sampled in the lower age ranges originally, then any regression-based scores derived will tend to perform more poorly in extreme score ranges. This could explain why 40% of postsecondary-aged students in the present sample were said to have FSIQ scores of 85 or less using Canadian norms, whereas if IQ scores are distributed normally one would predict that only about 16% of the population should have scores in this range. Whatever the cause, our data suggest that one must question the representativeness of the Canadian normative sample in the younger age ranges.

The results from the present study make it clear that clinicians will obtain very different scores for their clients depending on which normative system of the WAIS-IV they use. Furthermore, clinicians will obtain significantly lower scores using the Canadian normative system when testing students with FSIQs below the mean, and also for those under age 29. The score differences for the WAIS-IV are in fact greater than those reported for the WAIS-III by Iverson et al. (2006). Clinicians must therefore be mindful of this when choosing by which method an individual will be evaluated and must make this decision *a priori* rather than in rescoring after the fact in an effort to ensure qualification for a specific diagnosis or support program.

At this point, we would recommend that Canadian clinicians employ Canadian norms any time when all tests administered in the assessment battery have Canadian norms available or where tests were co-normed with a Canadian sample. In contrast, it would be important to employ American norms anytime that other tests in the assessment battery provide only American normative data or where co-norming data exists only for the American sample. Given the large discrepancy that is observed when scoring by means of these two different normative conversions, clinicians must focus on comparisons in a more homogeneous population especially when attempting to determine if any type of impairment or reduction in functioning has occurred. Another reason would be to use a prediction method for estimating past IQ, such as the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) or the Oklahoma Premorbid Intelligence Estimate-3 (OPIE-3; Schoenberg, Duff, Scott, & Adams, 2003) but see also Lange, Schoenberg, Saklofske, Woodward, and Brickell (2006). Some psychologists like to compare a specific patient's WAIS-IV scores to known clinical groups, either from the Technical Manual (Wechsler, 2008a) or from the literature. For example, the performances of mixed clinical samples (e.g., Chaudhry & Ready, 2012; Cullum & Larabee, 2010; Heyanka, Holster, & Golden, 2013; Nelson, Canivez, & Watkins, 2013), or groups of patients with schizophrenia (e.g., Michel et al. 2013), Alzheimer's disease and dementia (e.g., Cullum & Lacritz, 2009, 2012), neurotoxic exposure [e.g., Bolla, 2012], Chronic Pain [e.g., Greve, Bianchini, & Ord, 2012]), Autism Spectrum Disorder (e.g., Holdnack, Goldstein, & Drozdick, 2011), HIV (e.g., Manly et al. 1998), traumatic brain injuries (e.g., Brooks, Holdnack, & Iverson, 2011), multiple sclerosis (e.g., Ryan, Gontkovsky, Kreiner, & Tree, 2012), or attention-deficit/hyperactivity disorder (e.g., Goodwin, Gudjonsson, Sigurdsson, & Young, 2011) are available in the literature for the American norms. These are all good reasons to use the American norms. In general, if the WAIS-IV is used in isolation with a Canadian patient, or it is used with other tests that have Canadian norms, then the Canadian normative data may be appropriate. A caveat to this, however, is that additional research is needed on the sensitivity and specificity of the Canadiannormative data for identifying Intellectual Disability.

Funding

This work was supported in part by funding from the Ministry of Training, Colleges, and Universities (MTCU) of Ontario. The views expressed in this paper do not necessarily reflect those of MTCU.

Conflict of Interest

None declared.

References

Angoff, W. H., & Robertson, G. J. (1987). A procedure for standardizing individually administered tests, normed by age or grade level. *Applied Psychological Measurement*, *11*, 33–46.

Beal, A. L. (1988). Canadian content in the WISC-R: Bias or jingoism. *Canadian Journal of Behavioral Science*, *20*, 154–166.

Beal, A. L., Dumont, R., Cruse, C. L., & Branche, A. H. (1996). Practical implications of differences between the American and Canadian norms for WISC-III and a short form for children with learning disabilities. *Canadian Journal of School Psychology*, *12 (1)*, 7–14.

Bolla, K. I. (2012). Neurotoxic injury. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 281–301). New York: Oxford University Press.

Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment*, *18*, 156–167.

Brooks, B. L., Iverson, G. L., & Holdnack, J. A. (2013). Understanding and using multivariate base rates with the WAIS-IV/WMS-IV. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV/WMS-IV/ACS: Advanced clinical interpretation*. San Diego, CA: Elsevier Science.

Chaudhry, M., & Ready, R. (2012). Differential effects of test anxiety & stress on the WAIS-IV. *Journal of Young Investigators*, *24*, 60–66.

Cotton, S. M., Crewther, D. P., & Crewther, S. G. (2005). Measurement error: Implications for diagnosis and discrepancy-models of developmental dyslexia. *Dyslexia*, *11930*, 186–202.

Cullum, C. M., & Lacritz, L. H. (2009). Neuropsychological assessment in dementia. In M. F. Weiner, & A. M. Lipton (Eds.), *The American Psychiatric Publishing textbook of Alzheimer disease and other dementias* (pp. 85–103). Arlington, VA: American Psychiatric Publishing.

Cullum, C. M., & Lacritz, L. H. (2012). Neuropsychological assessment. In M. F. Weiner, & A. M. Lipton (Eds.), *Clinical manual of Alzheimer disease and other dementias* (pp. 65–88). Arlington, VA: American Psychiatric Publishing.

Cullum, C. M., & Larrabee, G. J. (2010). WAIS-IV use in neuropsychological assessment. In L. G. Weiss, D. H. Saklofske, D. Coalson, & S. E. Raiford (Eds.), *WAIS-IV clinical use and interpretation* (pp. 167–188). San Diego, CA: Academic Press.

Goodwin, E., Gudjonsson, G. H., Sigurdsson, J. F., & Young, S. (2011). The impact of ADHD symptoms on intelligence test achievement and speed of performance. *Personality and Individual Differences*, *50 (8)*, 1273–1277.

Gorsuch, R. L. (2003). Norming tests with the continuous parameter estimation model. Paper presented at the annual meeting of the American Psychological Association, Toronto, ON.

Greve, K. W., Bianchini, K. J., & Ord, J. S. (2012). The psychological assessment of persons with chronic pain. In G. J. Larrabee (Ed.), *forensic neuropsychology: A scientific approach* (2nd ed., pp. 302–335). New York: Oxford University Press.

Harrison, A. G., & Holmes, A. (2014). Mild intellectual disability at the postsecondary level: Results of a survey of disability service offices. *Exceptionality Education International*, *23 (1)*, 22–39.

Heyanka, D. J., Holster, J. L., & Golden, C. J. (2013). Intraindividual neuropsychological test variability in healthy individuals with high average intelligence and educational attainment. *International Journal of Neuroscience*, *123 (8)*, 526–531.

Holdnack, H. A. (2001). Wechsler Test of Adult Reading: WTAR. San Antonio: The Psychological Corporation.

Holdnack, J., Goldstein, G., & Drozdick, L. (2011). Social perception and WAIS-IV performance in adolescents and adults diagnosed with Asperger's syndrome and autism. *Assessment*, *18 (2)*, 192–200.

Holdnack, J. A., Schoenberg, M. R., Lange, R. T., & Iverson, G. L. (2013). Predicting premorbid ability for the WAIS-IV, WMS-IV, and WASI-II. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV/WMS-IV/ACS: Advanced clinical interpretation*. San Diego, CA: Elsevier Science.

Iverson, G. L., Lange, R. T., & Viljoen, H. (2006). Comparing the Canadian and American WAIS-III normative systems in inpatient neuropsychiatry and forensic psychiatry. *Canadian Journal of Behavioural Science*, *38 (4)*, 348–353.

Kahneman, D. (2011). *Thinking fast and slow*. Toronto, Canada: MacMillian.

Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues, (6th ed.)*. Belmont, CA: Thompson Wadsworth.

Lange, R. T., Schoenberg, M. R., Saklofske, D. H., Woodward, T. S., & Brickell, T. A. (2006). Expanding the WAIS-III estimate of premorbid ability for Canadians (EPAC). *Journal of Clinical and Experimental Neuropsychology*, *28 (5)*, 773–789.

Manly, J. J., Miller, S. W., Heaton, R. K., Byrd, D., Reilly, J., Velasquez, R. J., et al. The HIV Neruobehavioural Research Group (1998). The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. *Journal of the International Neuropsychological Society*, *4 (3)*, 291–302.

McDermott, P., Watkins, M., & Rhoad, A. (2014). Whose IQ is it?—Assessor bias variance in high-stakes psychological assessment. *Assessment*, *26 (1)*, 207–214.

Michel, N. M., Goldberg, J. O., Heinrichs, R. W., Miles, A. A., Ammari, N., & McDermid Vaz, S. (2013). WAIS-IV profile of cognition in schizophrenia. *Assessment*, *20 (4)*, 462–473.

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale–Fourth Edition with a clinical sample. *Psychological Assessment*, *25 (2)*, 618–630.

Ryan, J. J., Gontkovsky, S. T., Kreiner, D. S., & Tree, H. A. (2012). Wechsler Adult Intelligence Scale–Fourth Edition performance in relapsing–remitting multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *34 (6)*, 571–579.

Ryan, J., & Schnakenberg-Ott, S. (2003). Scoring Reliability on the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III). *Assessment*, *10*, 151–159.

Schoenberg, M. R., Duff, K., Scott, J. G., & Adams, R. L. (2003). An evaluation of the clinical utility of the OPIE-3 as an estimate of premorbid WAIS-III FSIQ. *The Clinical Neuropsychologist*, *17*, 308–321.

Wechsler, D. (1991). *The Wechsler intelligence scale for children-Third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1996). *WISC-III manual Canadian supplement*. Toronto: Harcourt Brace Canada.

Weschler, D. (2001). *Wechsler Adult Intelligence Scale-Third edition: Canadian technical manual*. Toronto, ON: Harcourt Canada.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth edition*. San Antonio, TX: Harcourt.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children–Fourth Edition: Canadian Technical Manual.* Toronto, Ontario, Canada: PsychCorp.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale–Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition: Canadian Technical Manual*. Toronto, Ontario: Pearson Canada.

Willis, J. O., & Dumont, R. (2006). And never the twain shall meet: Can response to intervention and cognitive assessment be reconciled? *Psychology in the Schools*, *43*, 901–908.