Journal of Psychoeducational Assessment
1–7

DOI: 10.1177/0734282915590064
jpa.sagepub.com
SAGE

Getting Back to the Main Point: A Reply to Miller et al.

Allyson G. Harrison[1], Alana Holmes[2], Robert Silvestri[2], and Irene T. Armstrong[1]

Abstract

Miller et al. have challenged the findings of our two previous studies, based largely on the assumption that our findings are biased due to the clinical sample used. However, they fail to address the primary tenet of our studies, namely, that clinicians will obtain different scores on the Wechsler Adult Intelligence Scale-IV (WAIS-IV) depending on whether Canadian or American norms are used. This reply seeks to provide empirical evidence supporting the existence of such score differences even when nonclinical samples are used, and identifies some of the clinical decisions that are potentially affected by choice of normative data.

Keywords

special education, educational testing, intelligence, Canadian norms, diagnosis, intellectual disability, learning disability, advice to practitioners, WAIS-IV

Getting Back to the Main Point

Miller et al. (2015) discuss many issues in their article, but do not address the primary tenet of our two studies (Harrison, Armstrong, Harrison, Lange, & Iverson, 2014; Harrison, Holmes, Silvestri, & Armstrong, 2015). It was not our intention to imply that the Canadian norms have no utility. Rather, our message is simple. When using the two sets of WAIS-IV norms, different scores will be obtained. This is important for Canadian clinicians because these professionals sometimes have to make a decision about which norms to use for very real reasons. As such, it is important for Canadian clinicians to understand the consequences of applying both sets of norms to make an informed decision in their clinical practices.

Normative data allow clinicians to determine how abnormal or deviant a score is relative to most other people in the general population. Such scores, in turn, are often used to determine whether a client meets published criteria for a clinical disorder or disability. Our data demonstrate that using Canadian norms will influence clinical conclusions in predictable ways in a specific subsample of the population, specifically in those below age 35 and those with below average Full Scale IQ (FSIQ) scores. Miller et al. (2015) actually acknowledge this latter point when they state,

[1]Queen's University, Kingston, Ontario, Canada
[2]Cambrian College, Sudbury, Ontario, Canada

Corresponding Author:
Allyson G. Harrison, Regional Assessment and Resource Centre, Queen's University, Mackintosh-Corry Hall, B100, 68 University Avenue, Kingston, Ontario, Canada K7L 3N6.
Email: harrisna@queensu.ca

> With a smaller SD and a large negative skew, the distribution of intelligence scores in the WAIS-IV Canadian sample becomes narrower and scores drop off more quickly at the lower tail relative to the U.S. curve. As a result, the difference between a typical and atypical ability case is greater using the Canadian norms than the U.S. norms. (p. 4)

This is precisely the issue with which Canadian clinicians must cope.

IQ and Eligibility for Programs and Services

Contrary to what Miller et al. (2015) suggest, clinicians frequently use norms to determine whether a client meets criteria for a diagnosis or for program eligibility. Even if the professional does not do this, many government-funded programs use cut-scores to determine program or funding eligibility, which means that IQ scores may indeed be used to make funding or service determinations. For instance, an environmental scan undertaken by the Nova Scotia Department of Community Services (2004) demonstrated that in most Canadian provinces, access to services and supports as a person with a developmental disability is determined, in large part, by FSIQ.

While we agree with Miller et al. (2015) that such diagnostic decisions should be based on more than just one IQ score on one day, the fact remains that IQ scores are still "routinely used to differentially classify mental disability" (McDermott, Watkins, & Rhoad, 2014, p. 207) and students in school boards across Ontario, Canada, are often assigned an exceptionality classification of "Mild Intellectual Disability (MID)" on the basis of IQ test scores alone (Harrison & Holmes, 2014). Furthermore, the College Council on Disability Issues (2009) and Ford (2014) stated that community colleges in Ontario have agreed that students identified as MID (a FSIQ between 71 and 78) or those with "borderline intellectual functioning" (FSIQ between 71 and 84) would qualify for academic accommodations and government-provided disability funding support of to up to $10,000 a year (c.f. Freeman, Harrison, & Holtermann, 2012). Harrison and Holmes (2014) identified that the majority of colleges in Ontario use IQ alone as the criterion by which their Disability Services Office (DSO) determines whether a student qualifies for either of these disability categories. Given our findings, this means that approximately 40% of students assessed in clinical settings would qualify for accommodations and government-funded bursary programs or disability tuition rebates if Canadian norms were used to calculate FSIQ scores, whereas only 17% would qualify using American norms.

FSIQ is also routinely used to determine giftedness and to identify IQ-achievement discrepancies as part of the diagnosis of a learning disability (Harrison & Holmes, 2012; Kozey & Siegel, 2008). It is also used to rule out intellectual ability problems as the primary cause of other conditions (e.g., Kamphaus, Worrell, & Harrison, 2005). Again, our data indicate that different percentages of individuals would qualify in these categories based on choice of norms.

Different Norms Produce Significantly Different Scores

Miller et al. (2015) suggest that our findings were due to both the nature of our clinical sample and also possible symptom exaggeration. These arguments, however, miss the point. It does not matter whether one compares scores from a clinical or nonclinical sample or whether good or poor effort was invested. The point is that use of different WAIS-IV normative data to interpret the same raw score produces different results. This issue has been demonstrated previously in the literature when applying different norms using the Controlled Oral Word Association Test (Iverson, Franzen, & Lovell, 1999) and the WAIS-III (Iverson, Lange, & Viljoen, 2006). In fact, the acquisition of different scores using the Canadian versus American norms can be shown quite simply without using clinical data. To demonstrate this, we carefully created a set of 20 raw scores that spanned the entire score range for each WAIS-IV subtest (e.g., raw scores were evenly

spaced from the bottom to the top end of the score range for each WAIS-IV subtest). We assumed only that Case 1 (lowest) would receive credit for all items up to the adult baseline question for each subtest, often about 3 points in total. Therefore, Case 1 obtained the lowest raw scores on each of the 10 core WAIS-IV subtests (assuming baseline items were passed), and the scores on each subtest increased equally for each subsequent case until Case 20 where the highest scores were obtained on each subtest. Next, we entered these raw scores into the WAIS-IV computer-scoring program (Scoring Assistant software), did not identify sex (an option in the scoring program), and gave all 20 entries the same birth date. We then calculated IQ, Index, and subtest scores using either the Canadian or American norms for each of the 20 cases. Finally, we altered the age of each case to correspond to one of the eight age classification ranges listed in the WAIS-IV normative manual (e.g., one set of 20 cases was scored using norms for 16- to 17-year-olds, then for 18- to 19-year-olds, 20- to 24-year-olds, 25- to 29-year-olds, 30- to 34-year-olds, 35- to 44-year-olds, 45- to 54-year-olds, and finally 55- to 64-year-olds) to investigate the score differences across various age groups over the spectrum of possible IQ scores.

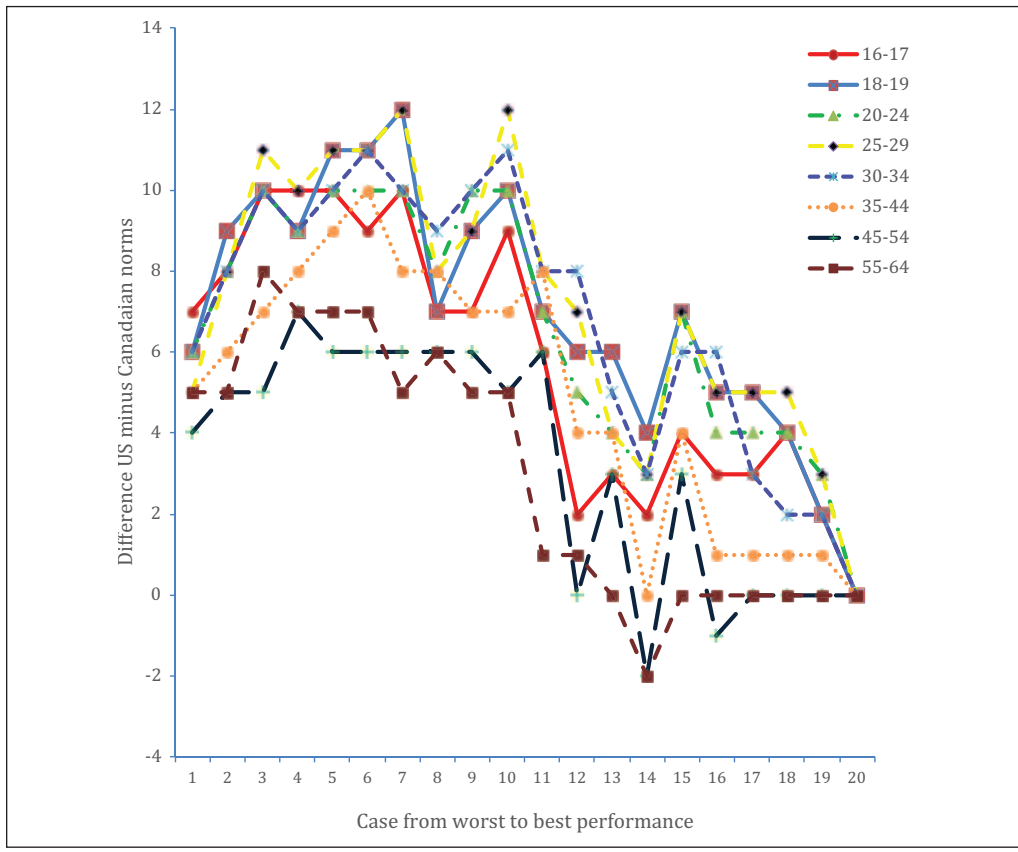

Figure 1. The difference between U.S. and Canadian-normed scores (positive when U.S.-normed scores are higher) as a function of age category for hypothetical scores from worst to best performance.

The main results are shown in Figures 1 and 2. As may be seen, one obtains quite different scores when using the two scoring systems regardless of clinical diagnosis. Indeed, the lower the

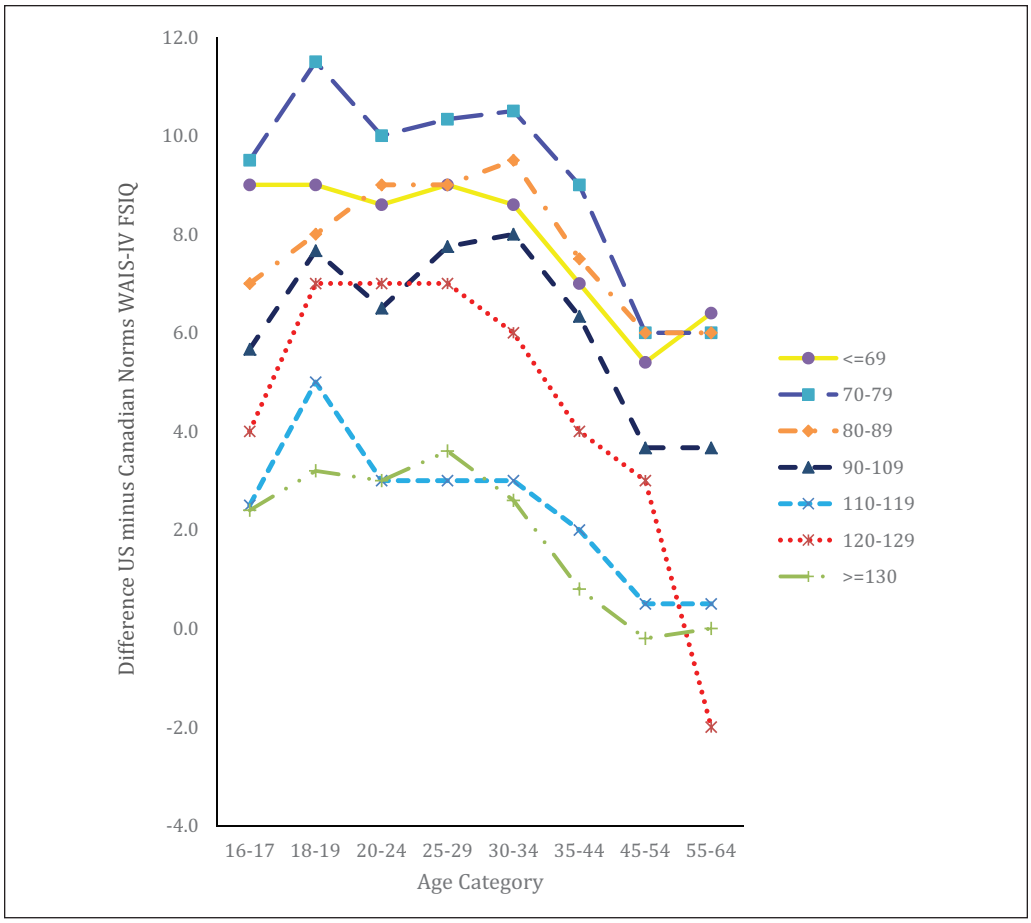

Figure 2. Mean difference in normed WAIS-IV FSIQ scores as a function of age and U.S. IQ category.
Note. FSIQ = Full Scale IQ.

raw score for an individual, the more likely a score difference of 10 to 12 IQ points on FSIQ will occur, but only in those below age 35.

When examining the FSIQ score differences, an ANOVA revealed a main effect for Age, $F(7, 144) = 14.4$, $p < .001$, and a main effect for IQ category, $F(6, 144) = 70.4$, $p < .001$,with no interaction present, $F(42, 144) = 0.6$, $p = .98$. Post hoc Tukey analysis of groups showed a significant effect for age: the oldest group (55-64) had the smallest mean differences compared with all other groups; the second oldest group (45-54) had smaller mean differences than all groups younger than 35. There were no mean differences among the youngest age groups (below 35), who all demonstrated the largest score differences ($\alpha = .05$).

As shown in Figure 2, score differences are systematically greater for lower IQ categories and generally lower at older age categories. When examining the effect of this score difference in different IQ ranges, post hoc Tukey analyses of the FSIQ categories show that the largest differences between normed scores fell in the lowest American-derived FSIQ categories e.g., 80-89, 70-79, 60-69) which were not different from each other. The highest FSIQ range (130+) and those in the 110 to 119 range returned the smallest mean differences, whereas those in the 120 to 130 range experienced a score change equal to those in the 90 to 109 range.

As with FSIQ differences, similar score differences were also found for General Abilities Index (GAI), and an ANOVA also showed a main effect for both Age, $F(7, 144) = 12.6$, $p < .001$,

and for GAI, $F(6, 144) = 49.6$, $p < .001$, with no interaction between the two factors, $F(42, 144) = 0.6$, $p = .927$. Post hoc Tukey comparisons showed that the oldest two age categories (45-64) had the smallest GAI differences but did not differ from each other, while the youngest age categories (16-34) did not differ from each other but showed significantly larger GAI differences than seen in the older age cohorts. Similarly, post hoc Tukey analyses of IQ categories showed the highest three IQ categories (110->130) to have the smallest GAI differences while the lowest three IQ categories (80-89, 70-79, 60-69) had the largest GAI differences; the youngest ages and lowest GAI scores have the largest differences in GAI.

Finally, in no case below age 35 did we obtain a higher FSIQ or GAI score when Canadian norms were used, even in the higher IQ ranges. Only in the age 45 and above categories did individuals in the highest IQ range obtain a higher Canadian score, but only by 1 or 2 points. In all but the most extreme case (i.e., perfect scores on all 10 subtests in which case all scores were identical), Canadian scores for individuals below age 35 were systematically lower than American-based scores. This demonstrates that, regardless of diagnostic status or level of effort invested, different scores are obtained when the different norms are used, and that this difference is greater in the lower IQ ranges and in the lower age ranges. These results are consistent with the findings presented in our two studies.

Discussion

The results from our studies 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 individuals with FSIQs below the mean, and also for those below age 35. Canadian clinicians require guidance on how to deal with this real difference rather than being distracted by other arguments. It is clear that the differences that arise depending on norms used have important implications, especially regarding their use in high stakes contexts such as disability determination.

A wide variety of questions could arise regarding how best to determine whether someone meets criteria for any disability or impairment. Who should make the judgment about which set of norms to use in high stakes assessments: individual psychologists, professional associations, DSOs, lawyers, or clients? What are clinicians to do when Canadian postsecondary students, referred because their high school provided them with an Individual Education Plan but no formal diagnosis, now obtain FSIQ and/or GAI scores that are below 75 using Canadian norms but significantly higher when using American ones? Do they now qualify as having a permanent intellectual disability even though our research suggests that 21% of clinic-referred students have such low scores using Canadian norms but only 4% when using American norms? Which norms should clinicians use when reassessing a Canadian student previously identified as having a learning disability who is seeking updated documentation and is planning to apply to both Canadian and American postsecondary institutions? Using Canadian IQ scores, this student has below average FSIQ and GAI scores that are equal to his or her academic achievement on a Canadian-normed achievement test, but using American norms, his or her IQ rises 12 points and is now Average and significantly different from his or her below average achievement scores (see Holmes & Rahemtulla, 2015, for evidence that Wechsler Individual Achievement Test–III [WIAT-III; Pearson, 2009] scores fail to change significantly depending on whether Canadian or American norms are used). Which is an accurate indication of this student's current level of functioning and disability status? If competing with American students, the data clearly indicate that this student meets diagnostic criteria for a learning disability, but using Canadian norms, the student would no longer qualify as there is no unexpected learning failure nor is their otherwise average thinking ability. Which scores, then, should be reported and used to make a disability determination? Questions of this sort require answers.

Apart from the WAIS-IV and the WIAT-III, no other commonly used psychoeducational or neuropsychological tests provide Canadian clinicians with Canadian norms. We must therefore agree strongly with the position statement posted by Pearson Canada (2014), which mirrors closely the advice provided in Harrison et al. (2014):

> The use of any psychological or cognitive test can, at times, yield results which appear to be inconsistent with the examinee's presentation or performance on other measures. For these reasons, practitioners making comparisons across test instruments regarding individual performance should use a consistent normative dataset. This means that if other instruments are being used for which Canadian norms are not available, then U.S. norms should be used to allow comparisons to be made to other tests also using U.S. norms. Similarly, if the purpose of assessment is to evaluate progress over time, practitioners may use U.S. tests and U.S. norms in order to compare performance against previous administrations using U.S. norms. Decisions regarding the normative set must be made a priori based on psychometric integrity and applicability to the target population. (p. 2)

Authors' Note

The views expressed in this article do not necessarily reflect those of the Ministry of Training, Colleges, and Universities.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded, in part, by the Ontario Ministry of Training, Colleges and Universities.

References

College Council on Disability Issues. (2009). *Mild Intellectual Disabilities (MID)*. Retrieved from http://www.disabilityissues.ca/english/Link_docs/MID_Guide.pdf

Ford, B. (2014, May). *Diagnosing eligibility for college accommodations: MID, BIF, LD and the DSM-5*. Presented at the College Committee on Disability Issues Conference, Toronto, Ontario, Canada.

Freeman, J. G., Harrison, A. G., & Holtermann, E. (2012). Navigating the postsecondary landscape. In B. Wong & D. L. Butler (Eds.), *Learning about learning disabilities* (4th ed., pp. 355-372). London, England: Elsevier.

Harrison, A. G., Armstrong, I. T., Harrison, L. E., Lange, R. T., & Iverson, G. L. (2014). Comparing Canadian and American normative scores on the Wechsler Adult Intelligence Scale–Fourth edition. *Archives of Clinical Neuropsychology*, *29*, 737-746.

Harrison, A. G., & Holmes, A. (2012). Easier said than done: Operationalizing the diagnosis of Learning Disability for use at the post-secondary level in Canada. *Canadian Journal of School Psychology*, *27*(1), 12-34. doi: 10.1177/0829573512437021

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.

Harrison, A. G., Holmes, A., Silvestri, R., & Armstrong, I. T. (2015). Implications for educational classification and psychological diagnoses when using the Wechsler Adult Intelligence Scale-Fourth Edition with Canadian as opposed to American norms. *Journal of Psychoeducational Assessment*, *33*(4), 299-311. doi:10.1177/0734282915573723

Holmes, A., & Rahemtulla, R. (2015, November). *Measuring the academic achievement gap between Americans and Canadians using the Wechsler Individual Achievement Test–Third Edition*. Poster accepted for presentation at the National Academy of Neuropsychology Annual Conference, Austin, TX.

Iverson, G., Franzen, M. D., & Lovell, M. R. (1999). Normative comparisons for the controlled oral word association test following acute traumatic brain injury. *The Clinical Neuropsychologist*, *13*, 437-441.

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

Kamphaus, R., Worrell, F. C., & Harrison, P. (2005). Principles for evaluation and eligibility determination for specific learning disabilities: A report of the ad hoc committee of division 16. *The School Psychologist*, *59*, 157-159.

Kozey, M., & Siegel, L. S. (2008). Definitions of learning disabilities in Canadian provinces and territories. *Canadian Psychology*, *49*, 167-171. doi:10.1037/0708-5591.49.2.162

McDermott, P., Watkins, M., & Rhoad, A. (2014). Whose IQ is it?—Assessor bias variance in high-stakes psychological assessment. *Assessment*, *26*, 207-214. doi:10.1037/a0034832

Miller, J. L., Weiss, L. G., Beal, L., Saklofske, D. H., Zhu, J., & Holdnack, J. A. (2015). Intelligent use of intelligence tests: Empirical and clinical support for Canadian WAIS-IV norms. *Journal of Psychoeducational Assessment*, *33*(4), 312-328. doi:10.1177/0734282915578577

Nova Scotia Department of Community Services. (2004). *Renewing the community support for adults program: A discussion paper*. Retrieved from http://www.novascotia.ca/coms/department/documents/Community_Supports_for_Adults_Renewal_Project_DiscussionPaper.pdf

Pearson. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Author.

Pearson Canada. (2014). *WAIS-IV* [Special note]. Toronto, Ontario, Canada: Author. Retrieved from http://www.pearsonassess.ca/content/dam/ani/clinicalassessments/ca/programs/pdfs/WAIS-IV_Special_Note_Dec2014.pdf