

# Comparison of the Wechsler Preschool and Primary Scale of Intelligence-Third Edition and the Leiter-R Intellectual Assessments for Clinic-Referred Children

Journal of Psychoeducational Assessment  
2022, Vol. 40(7) 825–838  
© The Author(s) 2022



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/07342829221105388  
[journals.sagepub.com/home/jpa](https://journals.sagepub.com/home/jpa)



Florence Renaud<sup>1,2</sup>, Marie-Julie Béliveau<sup>1,2</sup> ,  
Marc-Antoine Akzam-Ouellette<sup>1,2</sup>, Karine Jauvin<sup>1,2</sup>, and  
Fannie Labelle<sup>1,2</sup>

## Abstract

A review of clinical records was conducted for children with developmental, emotional, and behavioral difficulties who were assessed with both the Wechsler preschool and primary scale of intelligence-third edition (WPPSI-III<sup>CDN</sup>; Wechsler, 2004) and the Leiter international performance scale-revised (Leiter-R; Roid & Miller, 1997) within the same psychological evaluation. Forty children, ages 3–7, were included in this study. Pearson correlations showed that the IQ scores of the two instruments are strongly related ( $r > .70$ ;  $p < .001$ ). However, paired t-tests showed that overall Leiter-R scores ( $M = 99.03$ ) were significantly higher than WPPSI-III<sup>CDN</sup> scores (PIQ;  $M = 82.28$ , FSIQ;  $M = 75.24$ ) ( $p < .001$ ). The discrepancies between the instrument's scores were clinically important as the use of only one of the two instruments could result in misclassification of child intellectual ability. These results should prompt professionals working with this clinical population to be cautious when using results from a single instrument in a child's intellectual evaluation.

## Keywords

pre-school < participants, evaluation, intelligence/cognition, language

---

<sup>1</sup>Department of Psychology, Université de Montréal, Montreal, QC, Canada

<sup>2</sup>CIUSSS - Du Nord-de-l'île-de-Montréal, Montreal, QC, Canada

## Corresponding Author:

Marie-Julie Béliveau, Université de Montréal, Department of Psychology 90 av. Vincent-d'Indy, C-367 Montreal, QC H3C 3J7, Canada.

Email: [marie-julie.beliveau@umontreal.ca](mailto:marie-julie.beliveau@umontreal.ca)

There is no consensus on the definition of intelligence, but most agree that intelligence is the ability to understand, reason, react, learn, and adapt to environments (Legg & Hutter, 2007). The concept of intelligence is not unitary, as many forms of intelligence have been described (Lussier et al., 2018). Intellectual and cognitive tests are routinely administered to children in clinical settings as part of their overall psychological assessment. Intellectual and cognitive tests assess an individual's cognitive abilities by completing different tasks (Kaufman, 2018). These measures are necessary to provide diagnostic indicators, identify difficulties, determine the appropriate school placement, and support the implementation of a suitable intervention plan (Campbell et al., 2008; Schwan & Saklofske, 2005). The Cattell–Horn–Carroll theory (CHC; Schneider & McGrew, 2012) of intelligence is increasingly used to conceptualize intelligence in intellectual instruments (Kranzler et al., 2016). This theory postulates that intelligence is multidimensional and consists of different cognitive abilities arranged hierarchically. In this theory, stratum III's overarching general ability (*g*) comprises at least eight broad cognitive abilities in stratum II and over 80 narrow cognitive abilities in the last stratum (Kranzler et al., 2016). The *g* factor in CHC theory is represented by the intellectual quotient (IQ) in intellectual tests (Schneider & McGrew, 2012).

Different types of intellectual assessments are available for clinicians to use. Some instruments are unidimensional, for example, Ravens' Progressive Matrices (Raven & Court, 1998), and thus assess intelligence by the mean of one type of task, like matrix completion. However, most tests, for example, Wechsler intelligence scales (Wechsler, 2012), Stanford–Binet intelligence scales (Roid & Pomplun, 2012), Leiter intelligence scales (Roid & Miller, 1997), Universal nonverbal intelligence test (UNIT; McCallum, 2003), and Kaufman Assessment Battery for Children–Second Edition (K-ABC-II; Kaufman & Kaufman, 2013), are multidimensional and consider intelligence to be a combination of different abilities. These instruments can be interpreted under the CHC taxonomy. Some are explicitly based on CHC theory (e.g., K-ABC-II), and others are designed to comport with multiple intelligence models, including CHC (e.g., Wechsler Intelligence scales; Kranzler et al., 2016). In multidimensional assessments, it is essential to distinguish verbal and nonverbal measures (DeThorne & Schaefer, 2004). Some instruments, for example, Wechsler intelligence scales (Wechsler, 2012) and Stanford–Binet intelligence scales (Roid & Pomplun, 2012), consider language as a domain of intelligence. Thus, they include verbal subtests meant to capture language abilities. They also include other subtests designed to assess nonverbal abilities. However, these nonverbal subtests have verbal task instructions and sometimes require oral answers (DeThorne & Schaefer, 2004; McCallum, 2003, 2017). Language abilities are therefore necessary for successful task completion. In the present study, such instruments will be designated as verbal. Other cognitive instruments (e.g., Leiter-R, UNIT) try to eliminate the language bias by excluding any linguistic task and providing pantomime instructions for the children (DeThorne & Schaefer, 2004; McCallum, 2003). These assessments will be designated as nonverbal.

Completely nonverbal intellectual assessments are important since linguistic abilities could be problematic for many children. It could result in underestimation of their cognitive capacities because of language disorders, hearing problems, or other language barriers (Campbell et al., 2008; Mayes & Calhoun, 2003; McCallum, 2003). Co-occurrence of language problems has been documented as reaching rates of over 70% among children referred for emotional or behavioral difficulties (Benner et al., 2002; Smolla et al., 2015). These language impairments are often unnoticed by the parents or professionals surrounding the child because they are secondary to behavioral and emotional issues (Hollo et al., 2014). Children's problem behaviors are so salient that the adults surrounding the child often misperceive language difficulties as low intelligence, inattention, noncompliance, or defiance (Hollo et al., 2014). Thus, reliance on language abilities for assessing IQ in clinic-referred children may be problematic. The present study will examine IQ

scores documented via assessment of clinical records of previously assessed children with various developmental, behavioral, and emotional difficulties.

Intellectual assessments rely on standardized and validated instruments designed to give stable IQ scores to individuals. In the general population, IQ scores of different instruments are considered almost interchangeable for that reason (Grondhuis et al., 2018). On the other hand, the interchangeability between different intellectual instruments is not well documented for several clinical populations. Miller and Gilbert (2008) compared language-impaired children to typically developing peers on discrepancies between the Wechsler Intelligence Scale for Children-third edition (WISC-III; Wechsler, 1991) “nonverbal” scores and the UNIT (Bracken & McCallum, 1998) nonverbal scores. They found significant discrepancies between the scores for the clinical groups but not for the typically developing peers. The authors concluded that these discrepancies between instruments are important clinically and could result in misclassification or misdiagnosis (Miller & Gilbert, 2008). Similarly, studies with children diagnosed with an autism spectrum disorder (ASD) found discrepancies between the scores for the autistic group but no significant differences for the typically developing children group (Dawson et al., 2007; Nader et al., 2016). Autistic children showed higher scores on the Raven’s Progressive Matrices compared to the Wechsler intelligence scales (Dawson et al., 2007; Nader et al., 2016). Grondhuis and Mulick (2013) compared scores of the Leiter-R (Roid and Miller, 1997) and the Stanford–Binet fifth edition (SB5; Roid, 2003) for ASD children and found significantly higher scores on the Leiter-R with a mean discrepancy of 20.91 IQ points. Grondhuis et al. (2018) found similar results in an ASD children population, with average Leiter-R scores 9.6 IQ points higher than SB5 scores. In the 80s, a few studies were made comparing the IQ scores from Wechsler intelligence scales with the Leiter international performance scale (LIPS; Leiter & Arthur, 1940) for hearing impaired or deaf children and reported significant correlations (Boyd & Shapiro, 1986; Phelps & Branyan, 1988; Ulissi & Gibbins, 1984). Two studies found no significant differences between the scores of the Wechsler intelligence scales and the LIPS (Phelps & Branyan, 1988; Ulissi & Gibbins, 1984), while Boyd and Shapiro (1986) found significant discrepancies; LIPS mean scores were significantly higher than the WPPSI by an average of 9.95 IQ points (Boyd & Shapiro, 1986). A more recent study by Hickman (2007), comparing the overall Leiter-R and WISC-IV scores of children with moderate intellectual disabilities, reports a significant difference of six IQ points between the mean scores on the two scales with higher scores on the Leiter-R (Hickman, 2007). Studies conclude that in clinical populations, choice of an instrument can have a significant impact on the intellectual assessment (Baum et al., 2015; Miller & Gilbert, 2008; Motttron, 2004). Altogether, the association and differences between verbal and nonverbal intellectual assessments are not a largely studied subject, but it does seem that nonverbal assessments result in higher scores for clinical populations.

Including at least two intellectual instruments in the psychological assessment of a child could reduce assessment biases. Divergent results from the different instruments would suggest biases that might have impacted the results. That said, including several intellectual instruments can be costly and therefore superfluous in the case of convergent results. The most frequently administered intellectual instruments are the Wechsler scales (Kranzler et al., 2016). Thus, understanding the Wechsler scales results in convergences or divergences with less traditional instruments of intellectual ability is helpful in guiding a clinician’s choices while evaluating a child. To our knowledge, no study has compared IQ results of the Canadian version of the Wechsler Preschool and Primary Scale of Intelligence-third edition (WPPSI-III<sup>CDN</sup>; Wechsler, 2004) and the Leiter-R. These two instruments were widely used among clinicians working with preschool-aged children in Quebec (Béliveau et al., 2014) when the children of the present study were assessed. The Leiter-R had an important role in the nonverbal assessments of cognitive functioning in special education and psychology, notably with children having ASD or other

developmental disorders (McCallum, 2017; Roid et al., 2009). Although new versions of these instruments have been released since the evaluation of the children in the present study, understanding the convergence between the WPPSI-III<sup>CDN</sup> and the Leiter-R is helpful as newer versions of these instruments are similar to older versions in the structure and construct measured (Niileksela & Reynolds, 2019; Roid et al., 2013). Also, newer Wechsler intelligence scales are still verbal (Wechsler, 2012). Additionally, no study has assessed the correlation and differences between verbal intellectual assessments and nonverbal intellectual assessments for children referred to an external psychiatric clinic. Thus, the present study will determine if the addition of several intellectual instruments is relevant when assessing clinic-referred children by comparing the clinical results of these two instruments.

Thereby, the present study aims to explore the convergence between the IQ scores of a verbal instrument (WPPSI-III<sup>CDN</sup>) and a nonverbal instrument (Leiter-R) for clinic-referred children with emotional and behavioral difficulties. The convergence will be verified for the WPPSI-III<sup>CDN</sup> nonverbal IQ (Performance IQ) and full-scale IQ with the Brief IQ of the Leiter-R.

## Methodology

### Participants

Participants in the present study are young children from a large Canadian metropolitan area who were assessed in an early childhood psychiatric clinic by a trained psychologist between 2004 and 2015. The Research Ethics Board and the relevant administrative authorities authorized access to children's clinical records. Only children having a WPPSI-III<sup>CDN</sup> and a Leiter-R administered within the same intellectual assessment were included. These instruments were the appropriate versions at the time of the evaluations. Eighty-three children's clinical records meeting these inclusion criteria were found, of whom 43 had an incomplete or inconclusive assessment at one of the two intellectual instruments, resulting in missing IQ scores, and were therefore excluded. The final sample includes 40 children (70% male) having a nonverbal intellectual and full-scale score for the WPPSI-III<sup>CDN</sup> and the Leiter-R. Sociodemographic characteristics are presented in Table 1.

### Instruments

*Wechsler Preschool and Primary Scale of Intelligence-third edition (WPPSI-III<sup>CDN</sup>)*. The WPPSI-III is an individually administered IQ test for children aged two years six months to seven years three months (Wechsler, 2002). The test is divided into two age bands, the younger band covering the ages of 2:6–3:11 and the older band covering from 4:0 to 7:3. The WPPSI-III conceptualizes intelligence as a hierarchical structure with different specific abilities comprised in broad cognitive abilities. The conceptualization also postulates an underlying global aspect of intelligence (Wechsler, 2004). Initially, the WPPSI was developed without referring to theoretical foundations of intelligence, but the WPPSI-III was designed to tap more specific theoretically based abilities. That said, the WPPSI-III is not explicitly based on CHC theory, even if this instrument is strongly supported to measure a child's level of *g* (Lichtenberger & Kaufman, 2004). This instrument provides an overall estimate of IQ, called the full-scale IQ (FSIQ), and composite scores for the different subscales of specific domains of intelligence. For both age bands, there are scores for the Performance IQ (PIQ), Verbal IQ (VIQ), and General Language Composite (GLC). For the older age band, there is also a subscale for Processing Speed (PSQ) (Gordon, 2004; Wechsler, 2002). In the present study, the PIQ and the FSIQ are included. The PIQ is included in this study as an estimate of the nonverbal IQ as measured with the WPPSI-III, a verbal instrument. The VIQ will also be reported as a sample characteristic but will not be included in the analyses as it is not the

**Table 1.** Sociodemographic Characteristics of Participants.

Characteristic	<i>n</i>	%
Sex		
Female	12	30
Male	28	70
Parent		
Two parents born in Canada	10	27
One parent born in Canada	10	27
No parent born in Canada	17	46
Diagnosis <sup>a</sup>		
Communication disorder	34	85
Motor disorder	35	87.5
Disruptive disorder	16	40
Relational disorder	17	42.5
Mood disorder	1	2.5
Pervasive development disorder or intellectual disability	9	22.5
Other non-specified disorder	7	17.5
Versions of WPPSI-III <sup>CDN</sup>		
Younger age band (2:6–3:11)	14	35
Older age band (4:0–7:3)	26	65

Note. *N* = 40. Participants were on average 4 years 7 months old (*SD* = 1.1).

<sup>a</sup>Diagnostic categories are non-exclusive. On average, children belonged to three diagnostic groups.

focus of this study. The FSIQ comprises four core subtests for the younger age band and seven core subtests for the older age band. The PIQ comprises two subtests (Block design and Object Assembly) for the younger children and three subtests (Block design, Matrix Reasoning, and Picture concepts) for the older children. The PIQ measures fluid reasoning, spatial processing skills, attentiveness to detail, and visual-motor coordination skills (Wechsler, 2002). The PIQ taps into some of the broad abilities of CHC theory (fluid ability, crystallized ability, and visualization), but the PIQ subtests cannot be broken apart into these “pure” abilities (Lichtenberger & Kaufman, 2004). In the present study, the Canadian norms of the WPPSI-III were used (WPPSI-III<sup>CDN</sup>; Wechsler, 2004).

*Leiter international performance scale-revised (Leiter-R).* The Leiter-R is an individually administered IQ test for individuals’ ages 2–20 years 11 months. It is entirely nonverbal as neither the examiner nor the child must speak. The instructions are given in pantomime, and the child’s answers are motor (Roid & Miller, 1997). It is especially recommended for individuals with language deficits, hearing impairments, ASD, or non-native language speaking (McCallum, 2003; Roid & Miller, 1997). The Leiter-R was conceptualized based on hierarchical models of intelligence. This instrument was designed to measure several cognitive abilities described in CHC theory: fluid reasoning, visual–spatial ability, short-term memory, long-term retrieval, processing speed, and some aspects of crystallized general knowledge, as well as to measure a child’s *g* (Roid & Miller, 1997; Roid et al., 2009). The Leiter-R comprises 20 subtests divided equally into two test batteries: the visualization and reasoning (VR) and the attention and memory (AM) battery. The VR battery is used to estimate IQ scores. The Leiter-R offers two estimates of IQ, the full-scale IQ (FIQ) and the brief IQ (BIQ), which are intended as measures of *g* (Roid & Miller, 1997). For all ages, the BIQ is a shorter version of the FIQ and is composed of four subtests (Figure Ground,

Form Completion, Sequential Order, and Repeated Patterns). Only the BIQ was included in this study.

## Analyses

Data were analyzed using SPSS26 (IBM, Armonk, New York, United States). Descriptive analyses were run to characterize the sample. Pearson correlations were used to explore the association between the WPPSI-III<sup>CDN</sup> and Leiter-R scores. Paired t-tests were used to verify differences in IQ scores between the two instruments. The effect size was examined using Cohen's *d* (*d*). Following Cohen's (2013) guidelines and Sawilowsky's (2009) addition, small, medium, large, very large, and huge effect sizes will be reflected in the values of *d* equal to 0.2, 0.5, 0.8, 1.2, and 2.0, respectively (Cohen, 2013; Sawilowsky, 2009). Because FSIQ and PIQ scores were included for the WPPSI-III<sup>CDN</sup>, both scores were compared to the Leiter-R BIQ score. Thus, the statistical significance of the analysis that would have been set at  $p < .05$  was corrected at  $< .025$  using the Bonferroni method. Since not all children had results for the FSIQ of the WPPSI-III<sup>CDN</sup>, it will be explicitly noted if the *N* is lower than 40. T-tests and Kruskal–Wallis tests were done to explore possible sex, age, and cultural group differences in the PIQ, FSIQ, and BIQ scores, as well as the discrepancies between those scores. Nonparametric tests were used to assess cultural group differences as data does not distribute normally in each group. To control for multiple analyses, the significance was set at  $p < .017$  with the Bonferroni method. Age groups were created using the WPPSI-III<sup>CDN</sup> version used, accounting for assessment differences between the younger (2:6–3:11) and older (4:0–7:3) age band.

## Results

The mean IQ scores of the WPPSI-III<sup>CDN</sup> and the Leiter-R are presented in Table 2.

### Association

The Leiter-R score and the WPPSI-III<sup>CDN</sup> scores are significantly and strongly correlated (Cohen, 1988). Correlation between the Leiter-R BIQ and the WPPSI-III<sup>CDN</sup> PIQ is  $r = .77$  ( $p < .001$ ;  $n = 40$ ), similar to the FSIQ:  $r = .74$  ( $p < .001$ ;  $n = 29$ ). See Table 3.

### Difference

Paired t-tests also detected significant discrepancies between the Leiter-R and the WPPSI-III<sup>CDN</sup> scores. The mean difference between the BIQ score of the Leiter-R and the WPPSI-III<sup>CDN</sup> PIQ score is 16.75 IQ points ( $t(39) = 7.25$ ;  $p < .001$ ,  $d = 1.15$ ) which is considered a large effect size. The mean difference between the Leiter-R BIQ and the WPPSI-III<sup>CDN</sup> FSIQ is 28.31 IQ points ( $t(28) = 9.32$ ;  $p < .001$ ,  $d = 1.73$ ), which is a very large effect size. Overall, Leiter-R BQI scores are

**Table 2.** Mean IQ scores of the Leiter-R and the WPPSI-III<sup>CDN</sup>.

Measure		<i>n</i>	<i>M</i>	<i>SD</i>	Range
<b>WPPSI-III<sup>CDN</sup></b>	FSIQ	29	75.24	19.5	41–128
	PIQ	40	82.28	17.57	47–127
	VIQ	35	70.54	19.33	48–122
<b>Leiter-R</b>	BIQ	40	99.03	22.92	43–151

**Table 3.** Correlations between the Leiter-R and WPPSI-III<sup>CDN</sup> IQ scores.

Variable	<i>M</i>	<i>SD</i>	1	2	3
1. Leiter- R BIQ	99.03	22.92	-		
2. WPPSI-III <sup>CDN</sup> PIQ	82.28	17.57	.77**	-	
3. WPPSI-III <sup>CDN</sup> FSIQ	75.24	19.5	.74**	.89**	-

Note. \*\* $p < .001$ .

higher than WPPSI-III<sup>CDN</sup> PIQ and FSIQ scores. In Figure 1, the scatter plot of the individual's scores on the Leiter-R, and the WPPSI-III<sup>CDN</sup> are presented. The discrepancy between the two instruments' scores ranges from 0 to 63 IQ points.

### Sociodemographic group differences

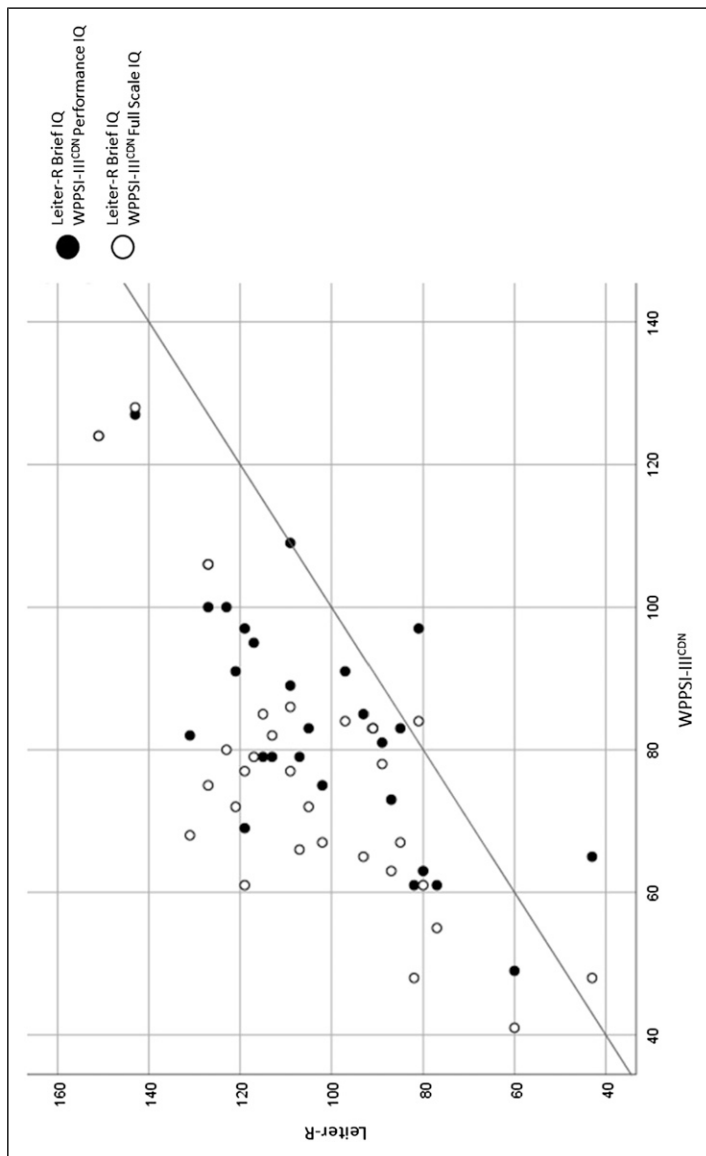
Results of t-tests suggest no differences between males and females for the mean scores on the Leiter-R ( $t(38) = -0.44, p = .665$ , males;  $M = 100.07, SD = 20.31$ , females;  $M = 96.58, SD = 29.02$ ) and the WPPSI-III<sup>CDN</sup> (PIQ;  $t(38) = -0.36, p = .724$ , males;  $M = 82.93, SD = 16.17$ , females;  $M = 80.75, SD = 21.19$  FSIQ;  $t(27) = -0.47, p = .642$ , males;  $M = 76.40, SD = 16.10$ , females;  $M = 72.67, SD = 26.55$ ) as well as no differences in discrepancies between the IQ scores of the two instruments (BIQ-PIQ;  $t(38) = -0.26, p = .799$ , BIQ-FSIQ;  $t(27) = -0.63, p = .537$ ).

Results of Kruskal–Wallis tests suggest no differences between children of parents both born in Canada, children with one parent born in Canada and children with both parents born outside of Canada for the Leiter-R ( $H(2) = 1.24, p = .538, M = 105.90, SD = 33.57, M = 100.00, SD = 20.48$  and  $M = 95.12, SD = 19.00$ ) and the WPPSI-III<sup>CDN</sup> (PIQ;  $H(2) = 0.86, p = .650, M = 89.30, SD = 24.60, M = 78.30, SD = 14.74$  and  $M = 80.18, SD = 15.53$ , FSIQ;  $H(2) = 2.53, p = .282, M = 85.63, SD = 31.33, M = 73.60, SD = 14.73$  and  $M = 69.60, SD = 7.55$ ) mean scores as well as the discrepancies between scores (BIQ-PIQ;  $H(2) = 0.86, p = .652$ , BIQ-FSIQ;  $H(2) = 2.82, p = .244$ ).

T-tests results show no significant ( $p > .017$ ) differences in WPPSI-III<sup>CDN</sup> scores (PIQ;  $t(38) = 2.43; p = .020$ , FSIQ;  $t(27) = 2.38; p = .025$ ) and discrepancies between instruments (BIQ-PIQ;  $t(38) = 1.86, p = .067$ , BIQ-FSIQ;  $t(27) = 1.47, p = .153$ ) between age groups. However, a significant difference in the BIQ score was found ( $t(38) = 3.26; p = .002$ ). The younger age group had significant higher Leiter-R BIQ scores ( $M = 113.43, SD = 21.04$ ) than the older age group ( $M = 91.27, SD = 20.27$ ).

### Discussion

The present study aimed to explore the convergence between two frequently used intellectual assessments, namely the Leiter-R and the WPPSI-III<sup>CDN</sup>. These two instruments rely on a hierarchical intelligence model and are intended to measure a child's *g*. That said, the instruments differ as the WPPSI-III<sup>CDN</sup> requires verbal knowledge while the Leiter-R is completely nonverbal. Pearson correlations showed a significant and strong association between the Leiter-R and the WPPSI-III<sup>CDN</sup> scores. However, despite this strong association between the two instruments, the mean scores obtained are statistically different for the two instruments. Overall, the Leiter-R has significantly higher scores than the WPPSI-III<sup>CDN</sup>. The mean discrepancy between the Leiter-R BIQ and the WPPSI-III<sup>CDN</sup> PIQ is 16.75 IQ points, and the mean discrepancy with the FSIQ is 28.31 IQ points. The Leiter-R and the WPPSI-III<sup>CDN</sup> are standardized with a mean of 100 and a 15 IQ point standard deviation. Score differences above the standard deviation will most likely



**Figure 1.** Distribution of the scores on the Leiter-R and the WPPSI-III<sup>CDN</sup>.

Note. Black dots represent the discrepancy between BIQ and PIQ, while white dots represent the discrepancy between FSQ and PIQ. The gray line represents perfect agreement between the scores of the two instruments. Dots above the line are individuals with higher scores on the Leiter-R. Dots below the line are individuals with higher scores on the WPPSI-III<sup>CDN</sup>.



change the functional descriptor (e.g., average, high, low, borderline) of a child's intellectual ability and is thus considered clinically meaningful.

The discrepancy between the Leiter-R BIQ scores and the WPPSI-III<sup>CDN</sup> FSIQ scores ( $D = 28.31$ ) is larger than with the WPPSI-III<sup>CDN</sup> PIQ scores ( $D = 16.75$ ). The larger difference with the FSIQ could be explained by the fact that this score explicitly considers language abilities to estimate the child's IQ. Knowing that in clinic-referred children with emotional or behavioral difficulties many have language impairments (Benner et al., 2002; Smolla et al., 2015), and considering that 85% of the children in this sample have a communication disorder, the FSIQ could underestimate the child's IQ. That said, the WPPSI-III<sup>CDN</sup> PIQ score, which is considered a nonverbal score, is still significantly lower than the Leiter-R BIQ. The PIQ score does not explicitly want to capture verbal abilities, but the administration is oral. Thus, children's language difficulties may impede their performance on the WPPSI-III<sup>CDN</sup> (Gallinat & Spaulding Tammie, 2014; Grondhuis & Mulick, 2013; Saar et al., 2018). Even so, the results of the present study and the current state of the scientific literature do not allow us to establish whether the WPPSI-III<sup>CDN</sup> underestimates the IQ or whether the Leiter-R overestimates children's IQ. Future research shedding light on this point is needed.

The present study's results are similar to past findings on the discrepancy of verbal and nonverbal intellectual assessments for clinical groups. Grondhuis and Mulick (2013) compared the Leiter-R to the Stanford–Binet for ASD children and found significantly higher Leiter-R scores by 22.43 points. The authors suggested that the Stanford–Binet and the Leiter-R do not capture the cognitive abilities of children having language deficits the same way as a partial explanation for the discrepancy between the IQ scores of the two instruments. Other studies also found similar discrepancies between IQ scores of verbal and nonverbal intellectual assessments for children of clinical groups (Boyd & Shapiro, 1986; Grondhuis et al., 2018; Hickman, 2007; Miller & Gilbert, 2008; Nader et al., 2016).

The effect sizes of the differences between the scores of the present study are really large (Cohen, 2013). This implies that inherent differences in the instruments mainly cause the differences observed in the IQ scores of the two instruments. The most significant difference between the Leiter-R and the WPPSI-III is that one is a nonverbal instrument while the other is a verbal instrument. As stated earlier, the verbal nature of the WPPSI-III might underestimate the children's IQ (DeThorne & Schaefer, 2004; McCallum, 2003, 2017). That said, it is still unclear to what extent verbal abilities influence nonverbal abilities. According to a recent study by Saar et al. (2018), with preschool-age children having language impairments, the nonverbal reasoning of these children evaluated with the WPPSI-III PIQ is lower than expected. The researchers of this study explain this finding by the language impairment of these children, either because the children do not adequately understand the verbal instructions of the WPPSI-III or because the children have delayed development of their inner speech, which could be necessary for nonverbal reasoning. Similarly, DeThorne and Schaefer (2004) suggest that verbal abilities influence nonverbal abilities because children can use verbal strategies to solve nonverbal problems. According to these researchers, this influence of verbal abilities on nonverbal abilities will always be present to some extent. Still, it can be minimized by using cognitive instruments that do not require any prior language knowledge. Thus, completely nonverbal instruments should be prioritized to assess a child with language impairments.

Another noticeable difference between the Leiter-R and the WPPSI-III that could partly explain the discrepancy between the scores of the present study concerns their differing motor demands. The Leiter-R requires simple motor demands to accommodate motor-impaired children (Roid & Miller, 1997). However, the WPPSI-III has complicated motor demands for some nonverbal subtests, such as block design included in both PIQ and FSIQ scores. Motor impairments often occur with language impairments (Hill, 2001; King-Dowling et al., 2015). In the present study,

87.5% of the children had motor impairments, mainly developmental coordination disorder. The higher motor demand on some of the subtests of the WPPSI-III could result in the underestimation of their intellectual capacities. Thus, the differences between the scores of the Leiter-R and the WPPSI-III could be due to the combined effect of language and motor skills of the children. In the present study, it cannot be ruled out that other inherent differences in the instruments play a role in the significant differences observed between the scores. Future studies should explore children's performance on the different subtests of the two instruments.

The instrument's norms could also partly explain the discrepancy between the IQ scores of the present study. The Leiter-R and the WPPSI-III<sup>CDN</sup> were standardized at a seven-year difference. IQ scores of these two instruments could thus have a two-point discrepancy explained by the Flynn effect (Flynn, 1984), as IQ increases three points per decade (Trahan et al., 2014). Also, American norms were used to assess the participants with the Leiter-R, and Canadian norms were used for the WPPSI-III. Higher performances are needed in Canadian norms for an equivalent IQ in American norms (Wechsler, 2004), especially in younger children. Canadian and American norms difference approximates three IQ points (Wechsler, 2004). Thus, assessing Canadian children with American norms on the Leiter-R could slightly overestimate their IQ. That said, norms alone cannot explain the magnitude of the differences in IQ scores obtained in the present study.

The discrepancies between the two instruments cannot be explained by age in the present study. That said, younger children had higher IQ scores than older children. Further research exploring age differences should be done, including broader age ranges and longitudinal data to verify IQ stability among very young clinical populations.

The cultural group was not a significant factor in IQ differences in the present study. However, the non-significant results could be due to a statistical power issue. Knowledge on IQ is mainly derived from research in Western countries, and past research has shown that the perception of intelligence varies between cultures (Sternberg et al., 2001). Cultural bias can affect intellectual assessments independently of language factors (Sternberg et al., 2001; van Wyhe et al., 2017). Further research should replicate these results with larger sample sizes.

### **Limitations**

Some limitations of the present study require discussion. First, the study sample may be biased by the clinical opinion of the professionals assessing the child. The evaluator may have administered the Leiter-R to the children they considered to "need" this kind of nonverbal instrument (Grondhuis & Mulick, 2013; Kuschner et al., 2007). Thus, the sample of the present study may be composed of children with difficulties that are not representative of all children consulting in child psychiatric clinics. This could result in an overestimation of the discrepancies between the instruments. Moreover, nearly half of the initially identified records had missing data and had to be excluded from the study. As the sampling method of the present study collects data in existing records, the data is limited to the information found in the records. Grondhuis and Mulick (2013) had a similar sampling method and found comparable findings (mean difference of 20.91 IQ points) to the present study. However, Grondhuis et al. (2018), who made their sample for research purposes, found smaller but still significant discrepancies (nearly 10 IQ points) between the Leiter-R and the SB5. That said, the Leiter-R is an instrument mainly used with individuals who have difficulties with language or for whom a more traditional way (e.g., Wechsler intelligence scales) of assessing IQ is not appropriate. Thus, understanding the discrepancies that may occur in clinical settings is helpful to professionals using these instruments. Furthermore, the present study included older versions of the instruments, probably no longer frequently used. That said, new versions of the instruments are similar to the older versions, and the present study results are thus relevant.

## Conclusion

The present study aimed to explore the associations and differences between two intellectual instruments, one verbal (WPPSI-III<sup>CDN</sup>) and the other completely nonverbal (Leiter-R), for children with developmental, emotional, or behavioral difficulties consulting in an early childhood psychiatric clinic. Results of the present study revealed significant differences in IQ scores between the two instruments, despite a strong correlation between children's scores. The discrepancy in scores reported in the present study could have broad clinical implications as using only one of the two instruments could result in misclassification of child intellectual ability. These results must be replicated with newer versions of the instruments. Further research is needed to understand the nature of the discrepancies between the IQ scores found in this study and validate that these discrepancies could be attributed to the children's verbal abilities. Studies using a randomized sampling method with a larger sample should be done to verify to what extent the current findings are generalizable to the entire clinical population. In clinical populations, including at least two intellectual instruments seems warranted before making a diagnostic or a clinical decision. Divergent results indicate measurement differences and can be interpreted with clinical judgment on a case-by-case basis. Several instruments, questionnaire results, and observations should guide a professional's decisions, even if it is time-consuming and more costly.

## Acknowledgments

This work would not have been possible without the support of the administrative authorities of the preschool child psychiatric clinic of the Rivières-des-Prairies psychiatric hospital. Special thanks are offered to Lucie Thibault, Roger Godbout, and Nicole Smolla, and to the clinical team for their incredible clinical work and collaboration: Julie Bélanger, Nathalie Valois, Chantale Breault, Dr Alain Lévesque, Dr Guylaine Gagné, Emmanuelle Cloutier, and Marie-France Le Lan.

## 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 project was partly funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

## ORCID iD

Marie-Julie Béliveau  <https://orcid.org/0000-0003-2370-0539>

## References

- Baum, K. T., Shear, P. K., Howe, S. R., & Bishop, S. L. (2015). A comparison of WISC-IV and SB-5 intelligence scores in adolescents with autism spectrum disorder. *Autism, 19*(6), 736–745. <https://doi.org/10.1177/1362361314554920>
- Béliveau, M.-J., Smolla, N., Breault, C., & Lévesque, A. (2014). L'évaluation développementale et cognitive de l'enfant d'âge préscolaire en clinique psychiatrique. [Preschool developmental and cognitive assessment tools in a psychiatric setting.]. *Revue Québécoise de Psychologie, 35*(1), 1–43.

- Benner, G. J., Nelson, J. R., & Epstein, M. H. (2002). Language skills of children with EBD: A literature review. *Journal of Emotional and Behavioral Disorders, 10*(1), 43–56. <https://doi.org/10.1177/106342660201000105>
- Boyd, J. & Shapiro, A. H. (1986). A comparison of the leiter-international performance scale to WPPSI performance with preschool deaf and hearing impaired children. *Journal of Rehabilitation of the Deaf, 20*(1), 23–26.
- Bracken, B. A. & McCallum, R. S. (1998). *Universal nonverbal intelligence test*. Riverside Publishing Company Chicago.
- Campbell, J. M., Brown, R. T., Cavanagh, S. E., Vess, S. F., & Segall, M. J. (2008). Evidence-based assessment of cognitive functioning in pediatric psychology. *Journal of Pediatric Psychology, 33*(9), 999–1014. <https://doi.org/10.1093/jpepsy/jsm138>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.*, 2nd edn (Erlbaum: ).
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Dawson, M., Soulières, I., Ann Gernsbacher, M., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science, 18*(8), 657–662. <https://doi.org/10.1111/j.1467-9280.2007.01954.x>
- DeThorne, L. S. & Schaefer, B. A. (2004). A guide to child nonverbal IQ measures. *American Journal of Speech-Language Pathology, 13*(4), 275–290. [https://doi.org/10.1044/1058-0360\(2004\)029](https://doi.org/10.1044/1058-0360(2004)029)
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51. <https://doi.org/10.1037/0033-2909.95.1.29>
- Gallinat, E. & Spaulding Tammie, J. (2014). Differences in the performance of children with specific Language Impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 57*(4), 1363–1382. [https://doi.org/10.1044/2014\\_JSLHR-L-12-0363](https://doi.org/10.1044/2014_JSLHR-L-12-0363)
- Gordon, B. (2004). Test review: Wechsler, D.(2002). The Wechsler preschool and primary scale of intelligence, (WPPSI-III). San Antonio, TX: The psychological corporation. *Canadian Journal of School Psychology, 19*(1–2), 205–220. <https://doi.org/10.1177/082957350401900111>
- Grondhuis, S. N., Lecavalier, L., Arnold, L. E., Handen, B. L., Scahill, L., McDougle, C. J., & Aman, M. G. (2018). Differences in verbal and nonverbal IQ test scores in children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 49*(1), 47–55. <https://doi.org/10.1016/j.rasd.2018.02.001>
- Grondhuis, S. N. & Mulick, J. A. (2013). Comparison of the Leiter International Performance Scale—revised and the Stanford-Binet Intelligence Scales. In: *Children with autism spectrum disorders American journal on intellectual and developmental disabilities*. 5th Edition, *118*(1), 44–54. <https://doi.org/10.1352/1944-7558-118.1.44>
- Hickman, W. F. (2007). *A comparison of students' performance on three cognitive measures and its implications for best practice* (Publication Number 3262427). The Johns Hopkins University]. ProQuest Dissertations & Theses Global. . <https://www.proquest.com/dissertations-theses/comparison-students-performance-on-three/docview/304869924/se-2?accountid=12543>
- Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders, 36*(2), 149–171. <https://doi.org/10.1080/13682820010019874>
- Hollo, A., Wehby, J. H., & Oliver, R. M. (2014). Unidentified language deficits in children with emotional and behavioral disorders: A meta-analysis. *Exceptional Children, 80*(2), 169–186. <https://doi.org/10.1177/001440291408000203>
- Kaufman, A. S. (2018). *Contemporary intellectual assessment: Theories, tests, and issues*. Guilford Publications.
- Kaufman, A. S. & Kaufman, N. L. (2013). *Kaufman assessment battery for children*. *Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals*.

- King-Dowling, S., Missiuna, C., Rodriguez, M. C., Greenway, M., & Cairney, J. (2015). Reprint of "Co-occurring motor, language and emotional-behavioral problems in children 3–6years of age". *Human Movement Science*, 42(1), 344–351. <https://doi.org/10.1016/j.humov.2015.06.005>
- Kranzler, J. H., Benson, N., & Floyd, R. G. (2016). Intellectual assessment of children and youth in the United States of America: Past, present, and future. *International Journal of School & Educational Psychology*, 4(4), 276–282. <https://doi.org/10.1080/21683603.2016.1166759>
- Kuschner, E. S., Bennetto, L., & Yost, K. (2007). Patterns of nonverbal cognitive functioning in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(5), 795–807. <https://doi.org/10.1007/s10803-006-0209-8>
- Legg, S. & Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and Applications*, 157(1), 17.
- Leiter, R. G. & Arthur, G. (1940). *Leiter international performance scale* (Vol. 1). Santa Barbara State College Press.
- Lichtenberger, E. O. & Kaufman, A. S. (2004). *Essentials of WPPSI-III assessment*. John Wiley & Sons.
- Lussier, F., Chevrier, E., & Gascon, L. (2018). Chapitre 2. Fonctionnement intellectuel. In *Neuropsychologie de l'enfant et de l'adolescent* (pp. 67–156). Dunod. <https://doi.org/10.3917/dunod.lussi.2018.01.0067>
- Mayes, S. D. & Calhoun, S. L. (2003). Analysis of WISC-III, Stanford-Binet:IV, and Academic Achievement test scores in children with autism. *Journal of Autism and Developmental Disorders*, 33(3), 329–341. <https://doi.org/10.1023/A:1024462719081>
- McCallum, R. S. (2003). *Handbook of nonverbal assessment* (Vol. 30). Springer.
- McCallum, R. S. (2017). *Handbook of nonverbal assessment- second edition*. Springer, . <https://doi.org/10.1007/978-3-319-50604-3>
- Miller, C. A. & Gilbert, E. (2008). Comparison of performance on two nonverbal intelligence tests by adolescents with and without language impairment. *Journal of Communication Disorders*, 41(4), 358–371. <https://doi.org/10.1016/j.jcomdis.2008.02.003>
- Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders*, 34(1), 19–27. <https://doi.org/10.1023/b:jadd.0000018070.88380.83>
- Nader, A.-M., Courchesne, V., Dawson, M., & Soulières, I. (2016). Does WISC-IV underestimate the intelligence of autistic children? *Journal of Autism and Developmental Disorders*, 46(5), 1582–1589. <https://doi.org/10.1007/s10803-014-2270-z>
- Niileksela, C. R. & Reynolds, M. R. (2019). Enduring the tests of age and time: Wechsler constructs across versions and revisions. *Intelligence*, 77(1), 101403. <https://doi.org/10.1016/j.intell.2019.101403>
- Phelps, L. & Branyan, B. J. (1988). Correlations among the Hiskey, K-Abc Nonverbal Scale, Leiter, and Wisc-r pErformance scale with public-school deaf children. *Journal of Psychoeducational Assessment*, 6(4), 354–358. <https://doi.org/10.1177/073428298800600404>
- Raven, J. C. & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759). Oxford psychologists Press.
- Roid, G. (2003). *Stanford-Binet Intelligence Scales—Fifth Edition*. Riverside Publishing.
- Roid, G. H. & Miller, L. J. (1997). *Leiter international performance scale-revised (Leiter-R)*. Stoelting, Vol. 10.
- Roid, G. H., Miller, L. J., & Koch, C. (2013). *Leiter international performance scale*. third edition. Stoelting Wood Dale.
- Roid, G. H. & Pomplun, M. (2012). The Stanford-Binet Intelligence Scales. Fifth Edition. In *Contemporary intellectual assessment: Theories, tests, and issues*, 3rd ed. (pp. 249–268). The Guilford Press.
- Roid, G. H., Pomplun, M., & Martin, J. J. (2009). Nonverbal intellectual and cognitive assessment with the Leiter International Performance Scale—Revised (Leiter-R). In *Practitioner's guide to assessing intelligence and achievement*. (pp. 265–290). John Wiley & Sons.

- Saar, V., Levänen, S., & Komulainen, E. (2018). Cognitive profiles of Finnish preschool children with expressive and receptive Language Impairment. *Journal of Speech, Language, and Hearing Research, 61*(2), 386–397. [https://doi.org/10.1044/2017\\_JSLHR-L-16-0365](https://doi.org/10.1044/2017_JSLHR-L-16-0365)
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 526–599. <https://doi.org/10.22237/jmasm/1257035100>
- Schneider, W. J. & McGrew, K. S. (2012). The cattell-horn-carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues*, 3rd ed. (pp. 99–144). The Guilford Press.
- Schwean, V. L. & Saklofske, D. H. (2005). 7 - assessment of attention deficit hyperactivity disorder with the WISC-IV. In A. Prifitera, L. G. Weiss, & D. H. Saklofske (Eds.), *WISC-IV clinical use and interpretation* (pp. 235–280). Academic Press. <https://doi.org/10.1016/B978-012564931-5/50008-6>
- Smolla, N., Bêliveau, M.-J., Noël, R., Breault, C., Lévesque, A., Berthiaume, C., & Martin, V. (2015). La pertinence de l'inquiétude parentale pour le développement langagier du jeune enfant référé en psychiatrie. [The relevance of parental worry for language development in clinic-referred preschoolers.]. *Revue Québécoise de Psychologie, 36*(3), 235–263.
- Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). *The predictive value of IQ*. Merrill-Palmer Quarterly (Vol. 1982, pp. 1–41).
- Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332–1360. <https://doi.org/10.1037/a0037173>
- Ulissi, S. M. & Gibbins, S. (1984). Use of the Leiter International Performance Scale and the Wechsler Intelligence Scale for children-revised with hearing-impaired children. *Diagnostique, 9*(3), 142–153. <https://doi.org/10.1177/073724778400900302>
- van Wyhe, K. S., van de Water, T., Boivin, M. J., Cotton, M. F., & Thomas, K. G. (2017). Cross-cultural assessment of HIV-associated cognitive impairment using the Kaufman assessment battery for children: A systematic review. *Journal of the International AIDS Society, 20*(1), 21412. <https://doi.org/10.7448/IAS.20.1.21412>
- Wechsler, D. (1991). *WMS-R: Échelle clinique de mémoire de Wechsler-Révisée: Mémoire figurative*. Editions du centre de psychologie appliquée.
- Wechsler, D. (2002). *WPPSI-III administration and scoring manual*. Psychological Corporation.
- Wechsler, D. (2004). *Wechsler preschool and primary scale of intelligence - Canadian Manual* (3 ed.). Harcourt Assessment.
- Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence—fourth edition*. The Psychological Corporation.