Applied Neuropsychology: Child

ISSN: 2162-2965 (Print) 2162-2973 (Online) Journal homepage: www.tandfonline.com/journals/hapc20

Processing speed and timed academic skills in children with learning problems

Benjamin J. Lovett, Allyson G. Harrison & Irene T. Armstrong

To cite this article: Benjamin J. Lovett, Allyson G. Harrison & Irene T. Armstrong (2022) Processing speed and timed academic skills in children with learning problems, Applied Neuropsychology: Child, 11:3, 320-327, DOI: 10.1080/21622965.2020.1824119

To link to this article: https://doi.org/10.1080/21622965.2020.1824119

Published online: 27 Sep 2020.

APPLIED NEUROPSYCHOLOGY: CHILD
2022, VOL. 11, NO. 3, 320–327
https://doi.org/10.1080/21622965.2020.1824119

Routledge
Taylor & Francis Group

Processing speed and timed academic skills in children with learning problems

Benjamin J. Lovett[a], Allyson G. Harrison[b], and Irene T. Armstrong[b]

[a]School Psychology, Teachers College, Columbia University, New York, New York, USA; [b]Regional Assessment and Resource Centre, Queen's University, Kingston, Canada

ABSTRACT
Information processing speed is commonly measured in intelligence and neuropsychological testing, and the scores from speed measures are considered in diagnostic and management recommendations for students with academic learning problems. However, this score usage often depends on assumptions about strong relationships between cognitive speed and the ability to perform actual academic tasks under time pressure. The primary purpose of the present study was to test the strength of these relationships empirically. In the present study, children with prior learning disability diagnoses (146 girls and 301 boys, ages 10–14 years old) completed diagnostic batteries that included measures of cognitive speed as well as timed academic skills. The relationships between the two types of measures were often modest (median $r = 0.25$), and the gap between processing speed and timed academic scores was typically approximately 1 standard deviation. The pattern of relationships suggested that superficial similarity in stimuli and task demands affected the strength of associations. These results suggest that timed academic skills cannot be reliably estimated based on processing speed scores, and there will often be significant gaps between the two. Therefore, making diagnostic judgments (e.g., learning disability diagnoses) or management recommendations (e.g., for extended time testing accommodations) should be based on more direct measures of relevant academic skills.

KEYWORDS
Academic skills; assessment; processing speed

Information processing speed (PS) has a long history and a distinguished place in psychology. Theories of intelligence often accord speed an important role in cognitive functioning; for instance, the current version of the Cattell–Horn–Carroll (CHC) Theory of Cognitive Abilities (Schneider & McGrew, 2018) names no fewer than four broad abilities related to speed (broad psychomotor speed, broad decision speed, broad cognitive speed, and retrieval fluency), with several narrow abilities under each of those. The earliest scientific measurements of individual differences in intelligence utilized reaction time measures (e.g., Cattell, 1890; Galton, 1883), and more recent research on the basic processes underlying intelligence has continued to explore speed (e.g., nerve conduction velocity, reaction time, etc.; Nettlebeck, 2011).

Most major intelligence tests measure PS as well. Virtually all of the cognitive assessment tools most commonly used (at least by school psychologists in the United States; see Benson et al., 2019) have subtests explicitly devoted to measuring speed. The Wechsler series of tests is by far the most commonly used; the version for most K-12 students, the Wechsler Intelligence Scale for Children (now in its fifth edition, the WISC-V; Wechsler, 2014), contains three such subtests, as well as a composite PS index score (the Processing Speed Index or PSI). The Woodcock-Johnson Tests of Cognitive Abilities (now in its fourth edition, the WJ-IV-COG; Schrank, McGrew, et al., 2014), which was developed to be aligned with CHC theory, also has several subtests and multiple composites involving speed.

The purpose of the present investigation is to examine the generality of PS scores with regard to one class of tasks: *timed measures of academic skills*. When children are assessed, whether in a clinical or a school setting, concerns over their ability to develop academic skills are often a primary issue. Even when children with known brain injury or dysfunction are assessed, the implications for academic functioning are typically of interest. The generalizability of neuropsychological test performance to real-world academic functioning has been much discussed (e.g., Chaytor & Schmitter-Edgecombe, 2003; Silverberg & Millis, 2009), and the present study was developed to evaluate a type of generalizability—from PS tasks that are, at least superficially, quite unlike most real-world activities, to timed academic tasks that appear to be very similar to many of those found in the child's naturalistic educational environment.

The interpretation and use of PS scores

Are measures of PS *expected* to generalize to more complex realistic tasks in various content domains? Despite their wide use in various assessment batteries, their interpretation

CONTACT Benjamin J. Lovett BL2799@tc.columbia.edu School Psychology, Teachers College, Columbia University, Box 120, 525 W. 120th St, New York, NY 10027, USA.

is not clear. On one hand, PS tasks tend to appear "clerical" in nature; they typically involve simple visual-motor tasks that emphasize checking and noting similarities and differences between visual stimuli. Consistent with this, Schneider (2013) described the term "processing speed" as misleading, since "it is not the speed at which all mental processing occurs" but instead "the speed and fluency at which very simple repetitive tasks … can be performed" (p. 312). Similarly, Beal et al. (2019) described the different index scores on the WISC-V as representing members of a team, and the PSI was "the clerk," a team member who "is not expected to make any really important decisions. The job of the clerk is to do what they are told to do, get it done fast, and not make any mistakes" (p. 32). These comments seem to suggest that PS is a low-level quality only related to intelligence in an adjunctive, supportive role.

On the other hand, at other times, PS is described as having wide implications for cognitive performance. Weiss et al. (2019) argued that "it would be a mistake to think of the [WISC-V] PSI as a measure of simple clerical functions that are not relevant or related to intellectual functioning" (p. 117), and Schneider (2013) also reported that PS has "considerable predictive utility" when academic and occupational outcomes are criterion variables. In a popular book for parents concerned about their children's low PS, Braaten and Willoughby (2014) argued for an even broader influence of PS deficits, as leading to interference with friendship formation and maintenance, and eventually to depression and anxiety. This second portrait, then, paints PS as having broad influence in a number of major life activities.

Despite their unclear meaning, PS scores are used regularly in making important clinical decisions. One use is in the diagnosis of learning disabilities (LD); one popular approach to diagnosis, the Pattern of Strengths and Weaknesses (PSW) approach, involves identifying an academic weakness and an underlying cognitive weakness, in the context of otherwise normal cognitive and academic functioning (e.g., Flanagan et al., 2010). The evidence that a cognitive weakness *underlies* an academic weakness comes from research on the relationships between cognitive and academic skills, and scholars who support the PSW approach argue that PS deficits relate to deficits in all three academic skill areas (reading, mathematics, and writing) (Flanagan et al., 2018). Therefore, a student's processing speed score could well be a factor in determining whether deficits in a number of academic areas would lead to an LD diagnosis, under the PSW approach.

A second common use of PS scores involves recommendations for educational services, particularly academic accommodations. Mather and Wendling (2018) suggested that students with low PS may need accommodations such as extended time, shortened assignments, and a reduction in activities involving copying. Braaten and Willoughby (2014) endorsed similar accommodations while also advocating for teaching students with PS deficits more about time and its management. Ofiesh et al. (2004) specifically mentioned PS scores from intelligence measures as one factor to consider when determining the need for extended time accommodations. Since such accommodations are becoming increasingly common and controversial (see e.g., Cahan et al., 2016; Lovett, 2020), it is important to determine whether PS deficits are actually a sound basis for recommending extra test-taking time.

In recent years, there has been significant debate about how much value measures of cognitive processes and abilities have for decisions about LD diagnosis and educational services. Response-to-Intervention (RTI; e.g., Fletcher & Miciak, 2018) models of LD diagnosis have become quite prevalent, and these models emphasize brief and direct measures of academic performance, to the (relative) exclusion of cognitive tests. Scholars coming from an RTI point of view have also argued that measures of academic skills provide more information than cognitive measures about what interventions are appropriate (see e.g., Burns et al., 2010).

The relevance of cognitive measures to accommodations decisions has been questioned as well. Recently, Lovett and Bizub (2019) provided a framework for determining accommodation needs; they argued for an emphasis on diagnostic measures that are as similar as possible to the real-world tasks for which accommodations are being considered. These scholars specifically mentioned a preference for diagnostic tests of *academic* skills rather than cognitive skills such as PS. Much of the issue of the generality of PS scores, then, involves their relation to actual academic skills.

The present study

The present exploratory investigation aimed to determine the strength of relationships (a) among PS measures, and (b) between PS measures and timed measures of *academic* skills, in a sample of middle school students who had been diagnosed with high-incidence developmental disorders such as learning and attention problems, and who were now being reevaluated. The sample was chosen because students with a history of academic problems would serve best to test the generality of processing speed when diagnoses of learning disabilities and educational services are being considered. We were interested in whether PS scores that are obtained *when evaluating a child with suspected learning problems* could be taken to indicate impairment (or lack of impairment) on timed tests of academic performance. As such, this sample was far more ecologically valid than one from the general population. (In real-world settings, students do not typically complete PS measures except as part of a referred evaluation.)

We hypothesized that the strength of the relationships observed would be determined by similarity in task stimuli and demands. For instance, we hypothesized that tasks involving abstract figural stimuli would be more strongly related to each other than to tasks involving stimuli with prior meaning (such as text, digits, etc.). However, given the exploratory nature of the study, and the fact that past research has generally involved general population samples rather than samples of students with learning problems, we did not make any a priori hypotheses about specific sizes of relationships.

Relatedly, we also calculated the mean absolute difference (MAD) between selected variables, particularly between PS scores and academic scores. At the level of an individual client, it is helpful to know how far apart the two types of scores are (see e.g., Floyd et al., 2008). For instance, if a student receives a PS score of 80, is it likely that their timed reading score will be within a few points of 80, suggesting a similar deficit in realistic academic activities? Lower MAD values typically correspond to higher positive correlations between variables, but they also reflect consistent mean differences between two measures, and can be more easily applied to the individual case.

Method

Participants

The participants were 447 students (146 girls, 301 boys) aged 10 to 14 years old ($M = 12.38$, $SD = 0.59$) from Southern Ontario who had been referred by their schools to participate in a comprehensive transition-to-high school program for those with LD offered by a regional assessment center; a component of this program was a reevaluation of cognitive and academic functioning. Previously, each of the referred students had been diagnosed with a learning disability (and sometimes a comorbid condition) prior to fifth grade. The initial learning disability diagnoses had been made in accordance with the Ontario Ministry of Education criteria (Ontario Public Service, 2017). Briefly, the Ontario Ministry of Education defines a learning disability as one of a number of neurodevelopmental disorders that persistently and significantly has a negative impact on a child's ability to learn and use academic and other skills. To receive this diagnosis, a child must have at least average ability for thinking and learning, demonstrate academic underachievement that is inconsistent with the intellectual abilities of the student, or show academic achievement that can be maintained only with extremely high levels of effort and/or additional support. Other potential causes for the academic underachievement (e.g., lack of exposure to appropriate education, hearing or visual dysfunction, low intelligence, another neurological or psychiatric disorder) are ruled out prior to making this diagnosis.

This project was carried out in a Canadian city with individuals primarily of middle socioeconomic status according to the Hollingshead index.

Procedure

All students were seen for psychoeducational assessments conducted or supervised by a licensed psychological service provider. Assessments were conducted over two or three sessions lasting two to three hours per session. Testing was done individually in a quiet room as per typical psychoeducational assessment protocol. Each assessment consisted of a battery of cognitive and achievement tests; however, there were differences in the exact composition of the battery, depending on the year in which the student was evaluated and each student's particular areas of concern. Therefore, the sample size for each measure varied; see Table 1. All research procedures were approved by the General Research Ethics Board at the institution where the data was collected.

Table 1. Descriptive statistics for sample.

Test	Subtest	*n*	*M*	*SD*
WISC IV	FSIQ	358	93.89	10.90
	VCI	337	98.75	12.18
	PRI	335	100.21	12.59
	WMI	336	89.35	11.03
	PS	336	88.15	13.63
	Sym Search	336	8.58	2.70
	Coding	335	7.14	2.82
TOWRE	Sight Word Eff	337	86.88	10.26
	Phonemic Decode	337	82.89	11.94
TOWL	Context Convent	79	6.96	2.33
	Context Language	79	8.01	2.46
	Story Construct	77	9.38	2.17
WJ (III & IV)	Reading Fluency	349	86.47	12.69
	Math Fluency	328	74.45	12.13
	Writing Fluency	299	87.28	13.07
WJ IV	GIA	74	89.77	10.78
WJ III/IV	Processing Speed	287	83.72	16.00
GORT	Rate SS	80	6.48	2.11
	Fluency SS	82	5.80	2.73
CTOPP	Rapid Naming SS	276	83.45	14.58

Note. WISC: Wechsler Intelligence Scale for Children; FSIQ: Full Scale IQ; VCI: verbal comprehension index; PRI: perceptual reasoning index; WMI: working memory index; PS: processing speed index; Sym Search: Symbol Search; TOWRE: Test of Word Reading Efficiency; TOWL: Test of Written Language; WJ: Woodcock Johnson; GORT: Gray Oral Reading Test; CTOPP: Comprehensive Test of Phonological Processing.

Measures

For the purposes of the study, selected scores from a set of instruments were recorded and entered into a data file. Certain scores came from the WISC-IV and the WJ-IV-COG or its previous edition (the WJ-III-COG), depending on when students were assessed. In addition, scores came from the current and previous editions of the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999), the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999), the Woodcock-Johnson IV Tests of Achievement (WJ-IV-ACH; Schrank, Mather, et al., 2014), the Gray Oral Reading Tests (GORT; Wiederholt & Bryant, 2012), the Wechsler Individual Achievement Test (WIAT; Wechsler, 2001), and the Test of Written Language (TOWL; Hammill & Larsen, 2009).[1] Below we discuss the selected tasks and scores in detail; we often use abbreviations that do not refer to the edition of different tests, since new editions or normative updates were published during the time period of the evaluations. All of the selected scores had good or excellent reliability (coefficients of 0.80 or above)[2] and validity evidence (typically concurrent and

[1]Generally, the scores and tasks from different editions of the same test were based on virtually identical stimuli. However, the WJ-IV-COG processing speed composite was based on different tasks in the 3rd and 4th editions.

[2]Owing to the archival nature of the study, we did not have access to test protocols for all participants, which would have allowed for sample-specific reliability analyses based on item-level data.

structural, based on relationships with other purported measures of the same constructs, as well as factor analyses), according to the tests' technical manuals. In addition, any score that was selected was available for at least 50 participants in the database.[3]

Processing speed scores

Subtest scores from three different test batteries that include measures of processing speed were included for analysis: the WISC, the WJ-COG, and the CTOPP.

The core processing speed subtests from the WISC (Coding and Symbol Search) were administered; scaled scores from these individual subtests as well as the resulting PSI composite score were recorded. Coding required students to draw abstract visual symbols associated with specific numbers according to a key, as quickly and accurately as possible (scoring is based on the number correctly copied within a 120 s time limit). In Symbol Search, students were asked to quickly and accurately compare a set of two abstract visual symbols on the left side of the page with five symbols presented on the right and determine if either of the two abstract symbols match one of the five options on the right side (scoring is based on the number correct in 120 s). Raw scores represent the total number of items completed in a 2-min time period. These are converted to scaled scores ($M = 10$, $SD = 3$), which are then combined to form the PSI (combining the scaled scores from these two subtests and producing a standardized score with $M = 100$, $SD = 15$).

PS composite scores from the WJ-IV-COG (called "Cognitive Processing Speed") and its prior (third) edition were recorded. The subtests making up these composites include such time-pressured tasks such as Decision Speed (in which students scan a series of pictured objects over four pages [10 rows per page] and circle the two pictures in each row that are conceptually related) and Visual Matching (in which students compare six numbers of increasing size [going from single to 3 digit numbers in a row; 60 rows in total] and circle the two that are identical). In both subtests, students are urged to work as quickly and as accurately as possible and are told they have 3 min to complete the task. For both tasks, total number correct is converted into a standardized score.

As a final PS measure, the Rapid Naming composite score from the CTOPP was recorded. This score is made up of performance on two tasks, Digit Naming Speed and Letter Naming Speed. In these tasks a series of digits or letters is presented on one page (9 digits/letters per row over 4 rows) and the individual is asked to name them aloud as quickly as possible. Time to complete each page is converted to scaled scores for each task and scaled scores from the two tasks are combined to form a Naming Speed composite score ($M = 100$, $SD = 15$). These tasks are thought to tap the retrieval fluency area of PS, and were hypothesized to be more closely related to timed reading skills than scores from PS scores that were based mostly on abstract figural stimuli.

[3]In the vast majority of cases, the bivariate sample sizes for our correlational analyses also had at least 50 participants; see below.

Timed reading scores

Two scores came from the TOWRE: Sight Word and Phonemic Decoding Efficiency. For both of these tasks, the score is the number of words read correctly within 45 s, converted to a scaled score. In the Sight Word Efficiency task, students read aloud a list of real but unrelated words as quickly as possible. In the Phonemic Decoding Efficiency task, the words one must read are nonsense words that require quick application of phonemic decoding rules (e.g., "zarp, bloink"). Two more scores came from the GORT, a Rate score (based on the speed at which students read increasingly complex passages aloud), and a Fluency score (based on the speed *and accuracy* of the oral reading of the passages). A final score came from the (Sentence) Reading Fluency task of the WJ-ACH. Here, students are told they have 3 min to read and answer as many simple sentences as possible. They read silently and indicate comprehension by answering yes or no on the page beside the sentence (e.g., a dog flies Y/N). Score is number of correct answers minus number of errors, converted to a standardized score.

Timed mathematics score

The only timed math score in the battery that met inclusion criteria was from the Mathematics Fluency subtest (later called Math Facts Fluency) on the WJ-ACH. On this paper-and-pencil task, students solve as many single-digit arithmetic problems as they can in 3 min (e.g., $4 + 1 = ?$). Score is number correct converted to a standardized score.

Timed writing scores

Two writing test scores were included in the present study. One score came from the Essay Composition task of the WIAT (in which students needed to write an elaborate essay under a strict 10 min time limit). The score is determined relative to a series of manualized rules for both content and word count. The second test came from the Writing Fluency subtest (later called Sentence Writing Fluency) of the WJ-ACH, in which students are told to quickly and accurately write as many simple sentences as possible using a series of unique item prompts. Score here is the number of correct sentences written in the time limit.

Three additional scores came from the Test of Written Language (TOWL) spontaneous writing section, where students are told to write a fictional story under a strict time limit, and the story is scored for Contextual Conventions (punctuation and spelling), Contextual Language (grammar and diction), and Story Construction (adequate plot, etc.).

Results

Table 1 presents descriptive statistics for our sample. On the vast majority of our measures, students' average scores were between the population mean (100 for standard scores, 10 for scaled scores) and 1 standard deviation below that value (85 or 7, respectively). The lowest sample mean scores were on the timed reading measures of the GORT, rate ($M = 6.48$, $SD = 2.11$) and fluency ($M = 5.80$, $SD = 2.73$).

Table 2. Intercorrelations of processing speed scores.

Test	WISC-IV Sym S r (n)	WISC-IV PS r (n)	WJ III/IV PS r (n)	CTOPP Rapid Naming r (n)
WISC-IV Coding	.56** (335)	.87** (335)	.55** (200)	.40** (256)
WISC-IV Sym S		.88** (336)	.49** (201)	.25** (257)
WISC-IV PS			.60** (201)	.36** (257)
WJ III/IV PS				.23** (156)

Note. $^{**}p < .01$. WISC: Weschler Intelligence Scale for Children; Sym S: Symbol Search; PS: processing speed composite; WJ: Woodcock Johnson; CTOPP: Comprehensive Test of Phonological Processing.

Table 3. Correlations between processing speed and timed reading scores.

Test	TOWRE SW r (n)	TOWRE PD r (n)	WJ Rdg fluency r (n)	GORT rate r (n)	GORT fluency r (n)
WISC-IV PS	.32** (317)	.10 (317)	.48** (258)	.30** (80)	.17 (80)
WJ PS	.17* (212)	.04 (212)	.25** (261)	−.03 (41)	−.11 (41)
CTOPP RN	.54** (254)	.48** (254)	.34** (200)	.36** (81)	.38** (81)
TOWRE SW		.69** (337)	.68** (271)	.85** (80)	.81** (80)
TOWRE PD			.50** (271)	.60** (80)	.73** (80)
WJ Rdg Fluency				.69** (49)	.57** (49)
GORT rate					.89** (82)

Note. $^{*}p < .05$; $^{**}p < .01$. WISC: Wechsler Intelligence Scale for Children; Sym S: Symbol Search; PS: processing speed composite; WJ: Woodcock Johnson; CTOPP: Comprehensive Test of Phonological Processing; TOWRE: Test of Word Reading Efficiency; SW: Sight Word; PD: Phonological Decoding; Rdg Fluency: Reading Fluency; GORT: Gray Oral Reading Test.

Table 4. Correlations between Processing Speed and Timed Writing Scores.

	WIAT Essay Comp r (n)	WJ III/IV Writing Fluency r (n)	TOWL Contextual Conventions r (n)	TOWL Contextual Language r (n)	TOWL Story Construction r (n)
WISC-IV PS	.46** (51)	.51** (212)	.21 (77)	.14 (77)	.12 (75)
WJ III/IV PS	.25 (51)	.43** (254)	.04 (54)	−.08 (54)	−.01 (54)
CTOPP RN	.27 (28)	.37** (153)	.24* (79)	.22 (79)	.19 (77)
WIAT EC		.57** (51)	– (0)	– (0)	– (0)
WJ III/IV Writ Flu			−.46 (36)	.39* (36)	.35* (36)
TOWL CC				.50** (79)	.36** (77)
TOWL CL					.48** (77)

Note. $^{*}p < .05$; $^{**}p < .01$; WISC: Wechsler Intelligence Scale for Children; PS: Processing speed; WJ: Woodcock Johnson; WIAT: Wechsler Individual Achievement Test; CTOPP: Comprehensive Test of Phonological Processing; RN: Rapid Naming; Writ Flu: Writing Fluency; TOWL: Test of Written Language; CC: Contextual Conventions; CL: Contextual Language.

Table 2 presents the intercorrelations of the PS scores. The intercorrelation of the two core WISC-IV subtests ($r = 0.56$) was similar to that of either subtest with the WJ-COG PS composite score (*r*s of .55 and .50), whereas all of these measures had lower correlations with the CTOPP rapid naming score (*r*s ranging from .23 to .40).

Table 3 presents the correlations between PS composite scores and timed reading scores. The WISC-IV PSI had higher correlations (*r*s ranging from .10 to .48) with each of the timed reading scores than did the WJ PS composite score (*r*s ranging from −.11 to .25). In addition, the CTOPP rapid naming score was almost always a stronger predictor of timed reading scores (*r*s ranging from .34 to .54) than were either of the other PS scores. The WISC and WJ PS scores had their highest correlations with the WJ Reading Fluency score, whereas the two highest correlations of the CTOPP rapid naming score were the TOWRE scores.

The single timed math score—the WJ math fluency score—was predicted best by the WISC-IV PSI, $r(240)$[4] = .52, $p < .01$, followed by the WJ PS composite score, $r(245) = .44$, $p < .01$, and least well by the CTOPP rapid naming score, $r(185) = .39$, $p < .01$.

Table 4 shows the correlations between PS and timed writing measures. All of the correlations predicting TOWL essay quality scores (by PS measures) were lower (*r*s ranging from −.08 to .24) than those predicting either the WIAT Essay Composition score (*r*s ranging from .27 to .53) or the WJ writing fluency score (*r*s ranging from .37 to .51).

Finally, Table 5 shows analyses of the absolute deviations between sets of scores from various measures of processing speed and academic fluency (derived from Tables 2 through 4). We calculated the absolute values of the differences (discrepancies) between participants' scores on various measures, and calculated the mean and standard deviation of those absolute differences (see e.g., Floyd et al., 2008). The mean absolute differences were between 10 and 19 standard score points; therefore, the typical participant would show about a 1 *SD* difference between their scores on any two PS measures, or between their scores on a PS measure and a timed academic measure.

Discussion

The purpose of the present study was to examine the strength of the relationship between PS measures and (a)

[4]Consistent with typical presentation of statistical results, the numbers in the parentheses here represent degrees of freedom (rather than sample sizes) for the statistical tests on the significance of *r*.

Table 5. Mean absolute differences between selected processing speed and academic measures.

Score 1	Score 2	Absolute difference *M* (*SD*)
WISC-IV PSI	WJ-III/IV PS composite	11.16 (9.0)
WISC-IV PSI	CTOPP Rapid Naming	13.31 (10.9)
WJ-III/IV PS composite	CTOPP Rapid Naming	15.52 (11.9)
WISC-IV PSI	WJ Reading Fluency	10.81 (8.3)
WJ-III/IV PS composite	WJ Reading Fluency	14.07 (10.6)
CTOPP Rapid Naming	WJ Reading Fluency	13.07 (10.1)
WISC-IV PSI	WJ Writing Fluency	10.22 (8.6)
WJ-III/IV PS composite	WJ Writing Fluency	12.78 (9.7)
CTOPP Rapid Naming	WJ Writing Fluency	13.64 (10.1)
WISC-IV PSI	WIAT Essay Composition	11.49 (8.3)
WJ-III/IV PS composite	WIAT Essay Composition	17.24 (14.1)
CTOPP Rapid Naming	WIAT Essay Composition	18.21 (12.5)

Note. WISC: Wechsler Intelligence Scale for Children; PS: Processing speed; WJ: Woodcock Johnson; CTOPP: Comprehensive Test of Phonological Processing; WIAT: Wechsler Individual Achievement Test.

other PS measures and (b) timed academic measures, all in a sample of younger adolescents who had been identified as having learning problems. We found that different abstract PS tasks correlated with each other at a median r of 0.56, but these abstract tasks correlated with rapid naming at a more modest level, median $r = 0.25$. PS composite scores showed highly variable relationships with timed academic measures, ranging from nonsignificant and negative (as low as $r = -0.11$) to as high as 0.54, with a median of 0.25.

Finally, superficially observable similarities in task demands and stimuli often appeared to account for the variability in correlations between PS and academic skill measures. The highest correlation found was between rapid oral naming of unrelated letters and digits (the CTOPP Rapid Naming Composite) and rapid oral reading of a list of unrelated words (the TOWRE Sight Word Efficiency score). More generally, rapid naming exhibited a stronger relationship with measures of timed reading than did PS for abstract stimuli. And with regard to timed writing tasks, PS measures predicted productivity of simple sentence writing fluency best on average, followed by prediction of performance on a timed essay, followed by prediction of grammatical and literary quality of a timed fictional story.

Substantive interpretation of correlations

Characterizing the strength of a correlation coefficient is a complex matter. A common strategy for interpretation is to consider the size in the context of other relationships in psychological research. Guidelines for these comparisons vary somewhat, but 0.56 is generally considered large or very large in this context, and 0.25 is considered to be medium or average relative to other psychological research (Hemphill, 2003; Funder & Ozer, 2019). However, although it is interesting that PS appears to be as strongly related to timed academic skills as are many variables in psychology to each other, this does not directly address our research questions. Appropriate interpretation of correlations depends on the theoretical framework that those correlations are testing (O'Grady, 1982).

In diagnostic contexts, different measures of PS are often used interchangeably (i.e., only one cognitive ability battery is given, and the PS score is interpreted in the same way regardless of the battery that it comes from). The theoretical framework supporting this practice is one in which one PS score could substitute for another, taking its place in any operational application of the framework. Rather than trying to estimate the size of an experimental effect, then, we are judging how similar two indices of the same construct are. Therefore, the standard of comparison is not the size of correlations found in other psychology research, but complete overlap or unity. The obtained correlation of 0.60 between the WISC and WJ PS composites suggests that the two scores are far from interchangeable; only 36% of the variance in one is explained by variability in the other.

The theoretical framework relating PS to timed academic skills varies depending on the use of the PS scores. In the case of LD diagnosis, the models that suggest use of PS scores do not specify how strong the relationship between PS and academic skills is thought to be (see e.g., Flanagan et al., 2018). However, in the case of accommodations planning, the PS scores are taken as indices of a student's ability to read test items, perform relevant calculations, or otherwise compose answers under time pressure. Therefore, the PS scores are again thought to be substitutes for something else (in this case, direct measures of timed academic skill performance), and so unity is again the appropriate comparison. The observed median correlation of 0.25 between PS composites and academic skill scores suggests that only about 6% of the variance in timed academic skills is typically explained by PS composite scores. Moreover, although in the case of these correlations, some relationships are much larger than others, even the highest correlation only showed about 29% overlap between a PS measure and a timed measure of academic skills.

Admittedly, the imperfect reliability of the measures employed will generally attenuate correlations somewhat (Goodwin & Leech, 2006), and so in some sense, the actual relationship between the underlying constructs of PS and timed academic skills is likely somewhat larger. However, for the operational use of PS measures, this is immaterial; clinicians must employ and interpret PS scores in their obtained, imperfect form, without access to what classical test theory refers to as "true scores." Therefore, our analyses show correlations that are to be expected in ecologically valid clinical settings.

Even the modest correlation coefficients potentially overestimate the "exchangeability" of two test scores (Floyd et al., 2008), since absolute differences between scores will not affect correlations if the rank-ordering of participants remains similar. We therefore calculated mean absolute differences (MADs) between different measures in our battery, and found that participants' scores typically varied by about 1 *SD* (between 10 and 18 standard score points) between any two measures. One *SD* is quite often the difference between average and below average skills; since our sample's mean PS scores were in the low average range, 1 *SD* above this would be average, whereas 1 *SD* below this would be quite low. A PS score in the low average range (as is common in students referred for learning problems) is therefore quite consistent with both severe deficits in timed academic skills (1 *SD* below the PS score) as well as perfectly average timed academic skills (1 *SD* above the PS score).

Theoretical implications of relationship patterns

We had not started this project with *precise* a priori hypotheses regarding which relationships between PS and timed academic skills would be stronger and weaker. We had only used a rough behavioral framework suggested by Lovett and Bizub (2019), where obvious differences in task demands would be expected to lower the strength of relationships. An alternative framework would use the CHC theory of cognitive abilities to explain (or hypothesize) relationships. Interestingly, the current version of CHC theory (Schneider & McGrew, 2018) lists simple timed academic skills (number facility, reading fluency, and writing fluency) as well as typical PS tasks (search/scanning and comparison/pattern recognition) as narrow abilities all under the same broad ability of *Gs*, broad cognitive speed. Most of our tasks would be measures of *Gs* under this rubric, and yet we found quite modest relationships between them. A few of our tasks, from the CTOPP and WIAT/TOWL, would appear to measure the narrow abilities of naming fluency and ideational fluency respectively, which CHC places under the broad ability of *Gr*, retrieval fluency. Our results do not appear to show lower correlations for tasks between *Gs*-*Gr* broad abilities than within each broad ability. However, we caution that a formal task coding manual using the most recent version of CHC theory was not available, and so these classifications are tentative. In addition, because all of our participants were children referred for learning problems, we cannot draw more general theoretical inferences about cognition-achievement relationships in the general population with the same confidence. If the relationships are stronger in the general population, that would better support the assumptions of the PSW model of learning disability diagnosis.

Clinical implications

One obvious clinical implication follows from the summary above: PS scores are a very poor index of timed academic skills (e.g., they average only 6% overlap in variance), and so if a student's levels of timed academic skills are an object of interest in a clinical evaluation, they should be measured directly rather than inferred on the basis of PS scores. Scores from intelligence tests are therefore generally an insufficient basis for conclusions about academic service needs, including accommodations on real-world tests of academic skills. Since learning disability evaluations sometimes fail to measure academic skills directly (Harrison et al., 2008), and recommendations of extended time accommodations are often made without specific supportive evidence from timed measures of academic skills (Weis et al., 2016), this suggests a need for changes in practice.

A second, more general clinical implication comes from the pattern and variability of the correlations between PS and academic skills measures: Clinicians should pay close attention to the precise features of task demands and stimuli when making inferences about relationships between diagnostic tests and real-world functioning. Academic skill measures will generally have more ecological validity (closeness to real-world academic tasks) than will PS measures, but even within those academic skills measures, some tasks will be far better than others at approximating real-world school activities. Admittedly, one role of diagnostic tests is to measure distinct skills and processes while controlling for the influence of others; neuropsychological tests have a long tradition of doing this carefully, and there will be times when psychoeducational assessments must do this as well. For instance, measuring mathematics calculation skills separate from reading skills cannot be done if a diagnostic task requires a student to read word problems before performing calculations to solve them. But depending on the inferences made on the basis of diagnostic evaluations, a desire for pure measurement of narrow skills must be weighed against ecological validity concerns, a tension that is acknowledged even in neuropsychological assessment (e.g., Howieson, 2019).

Limitations and directions for future research

This exploratory study had several expected limitations, each of which points to clear directions for future work. First, our analyses were limited by the assessment battery employed at a single evaluation center, and so we had a limited number of measures within each category. We would encourage additional research studying other measures of both PS and timed academic skills. If the same group of participants are given several measures each of PS and timed academic skills in the same domain (e.g., reading), multivariate statistical techniques could be especially helpful in showing the generalized relationship between the two constructs. Second, the strength of relationships between cognitive ability domains and academic skills sometimes varies across developmental level (see e.g., Flanagan et al., 2018), and so there may be different relationships between processing speed and timed academic skills in very young students who are first acquiring academic skills, or in postsecondary students who no longer share a common curriculum. Finally, given the use of diagnostic tests to determine accommodation needs, we would encourage work empirically relating both PS and timed academic skill measures to actual benefit from accommodations on real-world tests. Prior research has shown significant variability in benefit from extended time across different students (e.g., Elliott & Marquart, 2004; Spenceley & Wheeler, 2016), but relating benefit on actual classroom (or high-stakes standardized) tests to performance on diagnostic tests would allow for more targeted, effective accommodation recommendations.

References

Beal, A. L., Holdnack, J. A., Saklofske, D. H., & Weiss, L. G. (2016). Practical considerations in WISC-V interpretation and intervention.

In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (pp. 66–93). Academic Press.
Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology*, *72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004
Braaten, E., & Willoughby, B. (2014). *Bright kids who can't keep up*. Guilford.
Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *39*(1), 69–83.
Cahan, S., Nirel, R., & Alkoby, M. (2016). The extra-examination time granting policy: A reconceptualization. *Journal of Psychoeducational Assessment*, *34*(5), 461–472. https://doi.org/10.1177/0734282915616537
Cattell, J. M. (1890). Mental tests and measurements. Mind, *os-XV*(59), 373–391. https://doi.org/10.1093/mind/os-XV.59.373
Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–197. https://doi.org/10.1023/b:nerv.0000009483.91468.fb
Elliott, S. N., & Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children*, *70*(3), 349–367. https://doi.org/10.1177/001440290407000306
Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., McDonough, E. M., & Ortiz, S. O. (2018). Dual discrepancy/consistency operational definition of SLD. In V. C. Alfonso & D. P. Flanagan (Eds.), *Essentials for specific learning disability identification* (2nd ed., pp. 329–430). Wiley.
Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools*, *47*(7), 739–760. https://doi.org/10.1002/pits.20501
Fletcher, J. M., & Miciak, J. (2018). A response to intervention (RTI) approach to SLD identification. In V. C. Alfonso & D. P. Flanagan (Eds.), *Essentials for specific learning disability identification* (2nd ed., pp. 221–256). Wiley.
Floyd, R. G., Clark, M. H., & Shadish, W. R. (2008). The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice*, *39*(4), 414–423. https://doi.org/10.1037/0735-7028.39.4.414
Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202
Galton, F. (1883). *Inquiries into human faculty and its development*. Macmillan.
Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education*, *74*(3), 249–266. https://doi.org/10.3200/JEXE.74.3.249-266
Hammill, D. D., & Larsen, S. (2009). *Test of written language* (4th ed.). Pro-Ed.
Harrison, A. G., Nichols, E., & Larochette, A. C. (2008). Investigating the quality of learning disability documentation provided by students in higher education. *Canadian Journal of School Psychology*, *23*(2), 161–174. https://doi.org/10.1177/0829573507312051
Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *The American Psychologist*, *58*(1), 78–79. https://doi.org/10.1037/0003-066x.58.1.78
Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures. *The Clinical Neuropsychologist*, *33*(2), 200–208. https://doi.org/10.1080/13854046.2018.1552762
Lovett, B. J. (2020). Disability identification and educational accommodations: Lessons from the 2019 admissions scandal. *Educational Researcher*, *49*(2), 125–129. https://doi.org/10.3102/0013189X20902100
Lovett, B. J., & Bizub, A. L. (2019). Pinpointing disability accommodation needs: What evidence is most relevant? *Psychological Injury and Law*, *12*(1), 42–51. https://doi.org/10.1007/s12207-019-09341-1
Mather, N., & Wendling, B. J. (2018). Linking cognitive abilities to academic interventions for students with learning disabilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 777–809). Guilford.
Nettlebeck, T. (2011). Basic processes of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 371–393). Cambridge University Press.
Ofiesh, N. S., Hughes, C., & Scott, S. S. (2004). Extended test time and postsecondary students with learning disabilities: A model for decision making. *Learning Disabilities Research & Practice*, *19*(1), 57–70.
O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin*, *92*(3), 766–777. https://doi.org/10.1037/0033-2909.92.3.766
Ontario Public Service. (2017). *Special education in Ontario*. Available at: http://www.edu.gov.on.ca/eng/document/policy/os/onschools_2017e.pdf
Schneider, W. J. (2013). Principles of assessment of aptitude and achievement. In D. Saklofske, C. Reynolds, & V. Schwean (Eds.), *Oxford handbook of psychological assessment of children and adolescents* (pp. 286–330). Oxford University Press.
Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford.
Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Riverside.
Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Riverside.
Silverberg, N. D., & Millis, S. R. (2009). Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society: JINS*, *15*(1), 94–102. https://doi.org/10.1017/S1355617708090139
Spenceley, L. M., & Wheeler, S. (2016). The use of extended time by college students with disabilities. *Journal of Postsecondary Education and Disability*, *29*(2), 141–150.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Pro-Ed.
Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Pro-Ed.
Wechsler, D. (2001). *Wechsler Individual Achievement Test* (2nd ed.). Harcourt.
Wechsler, D. (2014). *Wechsler Intelligence Scale for Children-Fifth Edition*. NCS Pearson.
Weis, R., Dean, E. L., & Osborne, K. J. (2016). Accommodation decision making for postsecondary students with learning disabilities: Individually tailored or one size fits all? *Journal of Learning Disabilities*, *49*(5), 484–498. https://doi.org/10.1177/0022219414559648
Weiss, L. G., Holdnack, J. A., Saklofske, D. H., & Prifitera, A. (2019). Theoretical and clinical foundations of the Wechsler Intelligence Scale for Children-Fifth Edition. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V clinical use and interpretation* (2nd ed., pp. 97–128). Academic Press.
Wiederholt, J. L., & Bryant, B. R. (2012). *Gray Oral Reading Tests-Fifth Edition*: *Examiner's manual*. Pro-Ed.