

Reading accuracy and reading comprehension in autism: forms of assessment and implications for practice

Mark D Lauterbach and Katharine Pace Miles, Brooklyn, USA

Editorial comment

Studies have shown that many autistic students have good decoding skills but have greater difficulty in reading for meaning. Their scores on reading accuracy and reading comprehension tests are often very discrepant. In this paper, the authors consider the different types of reading assessment and report on a study which compared the test scores of able autistic students with those of students with learning disabilities. They suggest why scores might be discrepant in autism and comment on the suitability of different types of reading assessments for this group.

Address for correspondence

E-mail: MarkL@brooklyn.cuny.edu

Acknowledgements

We would like to thank the Cooke Center for Learning and Development for their support of this research.

Introduction

Reading is a complex task involving decoding skills and the ability to understand the text. Reading tests are designed to assess reading accuracy (decoding) or comprehension or both. Students with autism often have a different profile from that observed in typically developing readers (Atkinson, Slade, Powell and Levy, 2017; Bonifacci and Snowling, 2008; Gough and Hillinger, 1980; Torgeson, 2002). They often show large discrepancies between their word reading skills and their language comprehension skills with a tendency towards precocious (hyperlexic) word reading skills and poor comprehension (Ricketts, Jones, Happé and Charman, 2012; Saldaña, Carreiras and Frith, 2009).

Possible reasons underlying the difficulties in reading for meaning in autism

One theory is that problems in comprehension are caused by differences in social communication skills. Ricketts et al (2012) found a relationship between a child's level of difficulty in social understanding and their reading comprehension scores. Another related

theory is that difficulties with Theory of Mind, the ability to infer the mental states and knowledge of others, also leads to poorer comprehension scores (Atkinson, Slade, Powell and Levy, 2017). Other issues may affect the assessment of autistic students. Students may not be interested in the test materials; may be distracted by the social demands or sensory environment and may be slower to process instructions and materials (Arciuli, Stevens, Trembath and Simpson, 2013; Paynter, 2015).

Types of reading assessment

This study focuses on three types of assessment commonly used to assess a student's reading ability:

- Word Reading Assessments (WRAs) – reading accuracy
- Informal Reading Inventories (IRIs) – reading accuracy and comprehension
- Comprehensive Standardised Reading Assessments (CSRAs)

Word Reading Assessments

Word reading assessments (WRA) typically use graded lists of words that progress in difficulty. On some of these measures basal and ceiling scores are calculated based on how far the student is able to make it down the list while still reading accurately. Other versions calculate the total number of words read on the list either in a timed or untimed format. These scores are used to determine a student's 'reading level'. Because students are typically only looking at individual words presented in isolation, this test does not measure comprehension. However, it is argued that a student's own vocabulary and phonological awareness is likely to influence their performance on WRAs (Nation and Cocksey, 2009).

If a student is encountering a word for the first time in print, more advanced phonics and phonological skills will help the reader better decode the word or to come up with a close approximation of the pronunciation. If the word is in the student's spoken vocabulary s/he can match the approximation to an actual word and then produce an accurate reading of the word (Tunmer and Chapman, 2012). Additionally, Morris and his colleagues (2011) found a strong relationship between WRAs and oral reading fluency, which in turn, strongly supported reading comprehension.

Informal Reading Inventories (IRIs)

These assessments analyse the student's reading of connected text and measure both word reading accuracy and comprehension. Students read a passage that is judged to be at a particular reading level and then answer comprehension questions about the passage. Informal reading inventories (IRIs) have seen a surge in popularity as guided reading has become a central part of many reading curricula (Provost, Lambert and Babkie, 2010). If the student's reading accuracy (correct words read) and comprehension scores are high enough on a passage (what constitutes 'high' is based on a combination of the accuracy and comprehension score and is different for each test) then the student continues with the next passage, which is incrementally harder. This process continues until the student no longer scores above the acceptable level. The level of the last passage read is taken as the student's reading level. Many inventories have

gradations in their scoring between independent reading level (typically 95 per cent accuracy) and instructional reading level, 90 per cent accuracy (Provost, Lambert and Babkie, 2010).

Informal reading inventory assessments measure word reading and comprehension. Some have explicit vocabulary questions within the comprehension questions but they do not provide a direct measure of phonological or phonetic skills. While these letter skills are no longer important instructionally for more advanced readers, for early readers, these skills can be important in identifying underlying skill deficiencies. A key problem with these assessments is that text levelling is a difficult 'science' in which each programme develops its own holistic rubric that defines readability with little research supporting the outcomes (Hibbard, 2002). It is also confounded by the fact that the student's own experiences and knowledge outside of their reading skills can influence their comprehension scores (eg a student who is keen on baseball is likely to read a baseball passage at a higher level than a passage on another topic).

Comprehensive Standardised Reading Assessments (CSRAs)

These refer to assessments that measure a broad range of reading behaviours. These tests can cover anything from pre-reading skills such as phonological awareness, phonemic awareness, phonetic skills, concepts of print and alphabetic knowledge, to word reading, vocabulary and listening comprehension. Since what is relevant to the task of reading changes over the course of reading development (Ehri, 2014), a comprehensive assessment needs to either:

- target a particular range of reading development, eg the Nelson-Denny (Brown, Fishco and Hanna, 1993) which is only for high school and college level readers)
- have a wide range of content and have different start and end points based on basal and ceiling rules, eg the Woodcock Reading Mastery Test, third Edition (Woodcock, 2011) which covers kindergarten through high school)

or:

- have different materials for different abilities of readers and link the assessment psychometrically, eg the Group Reading Assessment and Diagnostic Evaluation (GRADE), which has test booklets for different grade levels from pre-kindergarten to high school, all psychometrically linked for continuous growth analysis (Williams, 2001).

CSRAs are better designed to accommodate autistic students and those with other educational needs as the material can be targeted, based on their reading ability (decoding/comprehension/vocab/phonological skills). However, it is crucial that the materials are matched appropriately to the student. A test that is too easy is uninformative, as is a test that is too hard (Crocker and Algina, 2008).

As with the other types of assessment, there are several potential problems to consider with this form of assessment. Firstly, the assessor needs to accurately match the student to the appropriate test level. Secondly, these tests typically provide an overall reading score. While most CSRAs have subsections based on different components of reading skill with corresponding scores, they typically provide a singular reading score (eg a grade equivalent (GE) and/or a standard score) based on the total score from the test. This works well for students whose reading skills are closely related. However, for students with uneven reading profiles, this may not give a 'true' measure of their reading skill. For example, their word reading score may be 6.0 GE and their comprehension score 3.0 GE, giving an overall reading score of 4.5 GE, but the student in question will not read like a typical student in the middle of 4th grade. Unless the teacher understood the breakdown of the different skills, it could lead to a misrepresentation of the child's reading ability.

Aim of the study

The purpose of this study was to explore whether autistic students have a discrepancy between their word reading scores and their comprehension scores, when assessed on the three types of assessment, as

compared to other students with special needs – in this case students with learning disabilities (LD), or mild/moderate intellectual disability delays (MID). The students with LD were defined as having one or more learning domains that were particularly low such as reading or maths but they were average in other domains. The students with mild or moderate intellectual ability had lower cognitive ability and had difficulties across many domains. The autistic students varied widely in terms of their intellectual ability and all were verbal. A panel of teachers and the educational director of the school categorised the students into these three categories; all were blind to the research hypothesis.

Methods

Participants

A total of 68 high school students from a private, urban school in the Northeast of the State participated in the study. They were enrolled in a special school that specifically educates students with special needs. There were 43 males and 25 females (see *Table 1* for details of the sample). Students were diagnosed with either a mild/moderate intellectual disability (MID), learning disability (LD), or autism. Students with little or no expressive verbal language were not included in this sample.

While information on race and socio-economic status (SES) was not provided for individual participants, data for the entire school population was available:

- 27 per cent of students were African
- 32 per cent American, Latino
- 6 per cent Asian or Pacific Islander
- 34 per cent white (non-Hispanic)
- 1 per cent multi-racial
(from parental self reports)

SES was not provided, but income levels were given with 38 per cent of families having an income below \$32,000, 21 per cent between \$32,000-70,000, 19 per cent between \$70,000–125,000 and 23 per cent of families with an income higher than \$125,000.

Table 1: Details of the participants

Sex	MID	LD	Autism	Total
Male	11	16	16	43
Female	11	12	2	25

MID=Moderate Intellectual Disability
LD=Learning Disability

Reading assessments used

Students were assessed on the three types of assessment discussed above, within the same two week period. The first measure was the Slossen Oral Reading Test (SORT-3), (Slossen and Nicholson, 2008). This is a series of graded reading lists, each consisting of 20 words that become progressively more difficult. Students continue reading the lists until they cannot read any of the words correctly on a given list. All correctly read words are counted to give the score. Internal consistency and test-retest stability all yield coefficients above .95.

The IRIs were based on levelled readers, running records, and comprehension questions provided by the levelled reading programme employed by the school (the Fountas and Pinnell Benchmark System (2011). Teachers administered the IRIs according to the standard protocols provided. They calculated two scores; the independent reading level (the text read at a 95 per cent accuracy rate with comprehension) and the instructional level (the text read at a 90 per cent accuracy rate with comprehension). The researchers used the reading level correlation tables provided by the test makers to convert the instructional level into a rough grade equivalent score for the analysis.

The Group Reading Assessment and the Diagnostic Evaluation (GRADE), (Williams, 2001) a traditional paper and pencil reading assessment, was also administered. The GRADE measures pre-reading skills for the pre-kindergarten and kindergarten levelled tests, with reading readiness and vocabulary added to the kindergarten levelled test. For all levels 1st grade and higher vocabulary, word reading and comprehension are assessed. Additionally, tests at each of the grade levels assess listening comprehension. Alphas for every test level are above .90.

Assessment procedure

All of the assessments were administered by school personnel as part of the mid year assessment period that is a regular part of the school's assessment programme. Standard protocols were followed for administering the different assessments. The head or assistant teacher administered the SORT-3 and the IRIs while the GRADE was administered in small groups by various school personnel trained on the assessment. Scoring for the SORT-3 and the IRIs was done by classroom personnel while staff in the assessment office scored the GRADE. The individuals conducting the assessments were not told of the research hypothesis until after the assessment period to minimise the potential influence on scoring. Data collection and conversion of IRI scores into grade levels was done by the lead researcher (ML).

Results

Grade equivalent (GE) scores were calculated for the SORT-3 and the GRADE. A measure of internal deviation (ID), similar to a standard deviation but done for each student, was calculated to measure the spread of each student's test scores. Results indicated that while the correlation between these three measures was relatively high (see *Table 2*), the SORT-3 was consistently the highest estimator of reading ability (GE = 4.8), while the IRI was the lowest (GE = 2.9) with the GRADE in the middle (GE = 3.5). The internal deviation was most pronounced in the group of students diagnosed with autism as compared to those diagnosed with LD or MID (see *Table 3* and *Figure 1*).

Table 2: Correlations of the scores on the three reading measures

Group	SORT-3	GRADE	IRI
SORT-3	1		
GRADE	.76**	1	
IRI	.70**	.84**	1

* P < .01

SORT-3 = Slossen Oral Reading Test, third edition

GRADE = Group Reading Assessment and Diagnostic Evaluation

IRI = Informal Reading Inventory

Table 3: Means and standard deviations of the three reading assessments for each diagnostic group

Group	SORT-3	GRADE	IRI	Average Score	Internal Deviation*
MID N=25	2.4 (1.96)	2.0 (1.26)	1.4 (0.79)	1.8 (1.12)	1.2 (0.79)
LD N=21	5.5 (2.31)	4.4 (1.74)	4.5 (1.71)	4.9 (1.43)	1.5 (0.89)
Autism N=22	6.4 (3.30)	4.0 (2.22)	2.8 (1.52)	4.4 (2.12)	3.0 (1.46)

ANOVA F 2, 65 = 18.06, p < .001, post hoc ASD > LD = MID
 MID = Moderate intellectual disability
 LD = Learning Disability
 SORT-3 = Slossen Oral Reading Test, third edition
 GRADE = Group Reading Assessment and Diagnostic Evaluation
 IRI= Informal Reading Inventory

Each student was categorised into one of six potential result patterns (see Table 4). The most common pattern was to have the lowest score on the IRI and the highest score on the word reading assessment (52 per cent of the total sample). However, the percentage with that pattern differed between the diagnostic groups, from 32 per cent for the MCI group to 82 per cent of the autistic group. While the Chi Square indicates that there is a statistically significant difference in the pattern of results, it does not show between which groups the difference lies. However, by partitioning the cross-tabulation table and analysing the various sections as suggested by Rindskopf (1996), key differences can be identified. This technique is similar to planned comparisons in an ANOVA (analysis of variance). Based on the research question, two partitions were performed. First, all the other patterns except for patterns one and two were compared (c2 = 10.88, P = .21) and the difference was not significant. The second partition was to compare the results pattern of the MID group to that of the LD group and they were not significantly different (c2 = 7.67, P = .18). This reduces the table to a 2x2 table (see Table 5) which makes it clear that the difference is found between the autistic group and the other groups.

Table 4: Percentage of students in each diagnostic group who showed each of the six score patterns

	Diagnosis	MID	LD	Autism	Total
Pattern	Pattern 1	8 (32%)	9 (41%)	18 (82%)	35 (52%)
	Pattern 2	5 (20%)	2 (9%)	2 (9%)	9 (13%)
	Pattern 3	4 (16%)	5 (24%)	0 (0%)	9 (13%)
	Pattern 4	1 (4%)	3 (14%)	0 (0%)	4 (6%)
	Pattern 5	0 (0%)	1 (5%)	0 (0%)	1 (2%)
	Pattern 6	7 (28%)	1 (5%)	2 (9%)	10 (15%)
Total		25	21	22	68

$\chi^2 = 23.37, P < .01$
 Pattern 1 = IRI < GRADE < SORT-3
 Pattern 2= IRI < SORT-3 < GRADE
 Pattern 3= GRADE < IRI < SORT-3
 Pattern 4= GRADE < SORT-3 < IRI
 Pattern 5= SORT-3 < GRADE < IRI
 Pattern 6= SORT-3 < IRI < GRADE

Table 5: Percentage of autistic students and students in the other two diagnostic groups combined who showed Score Pattern 1

	Diagnosis	All others	Autism	Total
Pattern	Pattern 1	17 (37%)	18 (82%)	35 (52%)
	All others	29 (63%)	4 (18%)	33 (49%)
Total		46	22	68

$\chi^2 = 11.99, P < .01$

To control for the fact that lower level test forms of the GRADE include foundational skills that could account for the higher scores over the IRIs, a second analysis was done using only test forms above grade 4. These only assess comprehension and vocabulary skills. The same pattern was found but is more pronounced with all of the autistic students having the IRI as their lowest score and only 47 per cent of the other students (in this case all LD) having that scoring pattern (see Table 6 for details).

Table 6: Percentage of autistic students and students in the other two diagnostic groups combined who showed Score Pattern 1, for the 27 students with 4th grade level assessment or higher on the GRADE

	Diagnosis	All others	Autism	Total
Pattern	Pattern 1	17 (37%)	18 (82%)	35 (52%)
	All others	29 (63%)	4 (18%)	33 (49%)
Total		46	22	68

$$\chi^2 = 11.99, P < .01$$

Discussion

This study supports previous findings that there is often a dramatic difference between word reading and comprehension skills for autistic students. The type of assessment used to assess reading comprehension was also found to be important. In looking at average scores and in the pattern of results, autistic students had significantly more difficulty with the IRIs than with the standardised reading measure, even when both were only measuring comprehension.

The discrepancy between reading accuracy and reading comprehension makes sense within the context of autism. Among the characteristics typically displayed by autistic students are difficulties with expressive and receptive language, figurative language, making inferences, reading social cues and social understanding in general. Given these issues, coupled with an apparent lack of impairment in the grapho-phonemic system, it makes theoretical sense that the SORT-3 was the assessment that the autistic students in this study typically performed best on. This test taps the least linguistically complex reading skills of the three assessments, and involves a minimum amount of social interaction between the test taker and the assessor. This is in contrast to the IRI, which was clearly the most challenging for all students but particularly challenging for the autistic students. The nature of the assessment highlights some of the difficulties typically found in

autism. After reading the text, students are asked to answer questions that often require making inferences and interpreting figurative language from the text within the context of a social interaction with the assessor.

While the third assessment, the GRADE, minimises the social interaction with the assessor, there is still potential for having to interpret figurative language and make inferences in the comprehension sections of the test. While the reading task demands are similar, albeit with multiple choice questions as opposed to open ended questions, the lack of social interaction, particularly the need for any verbal response from the test taker, makes the test demands easier for an autistic student.

Concluding comments

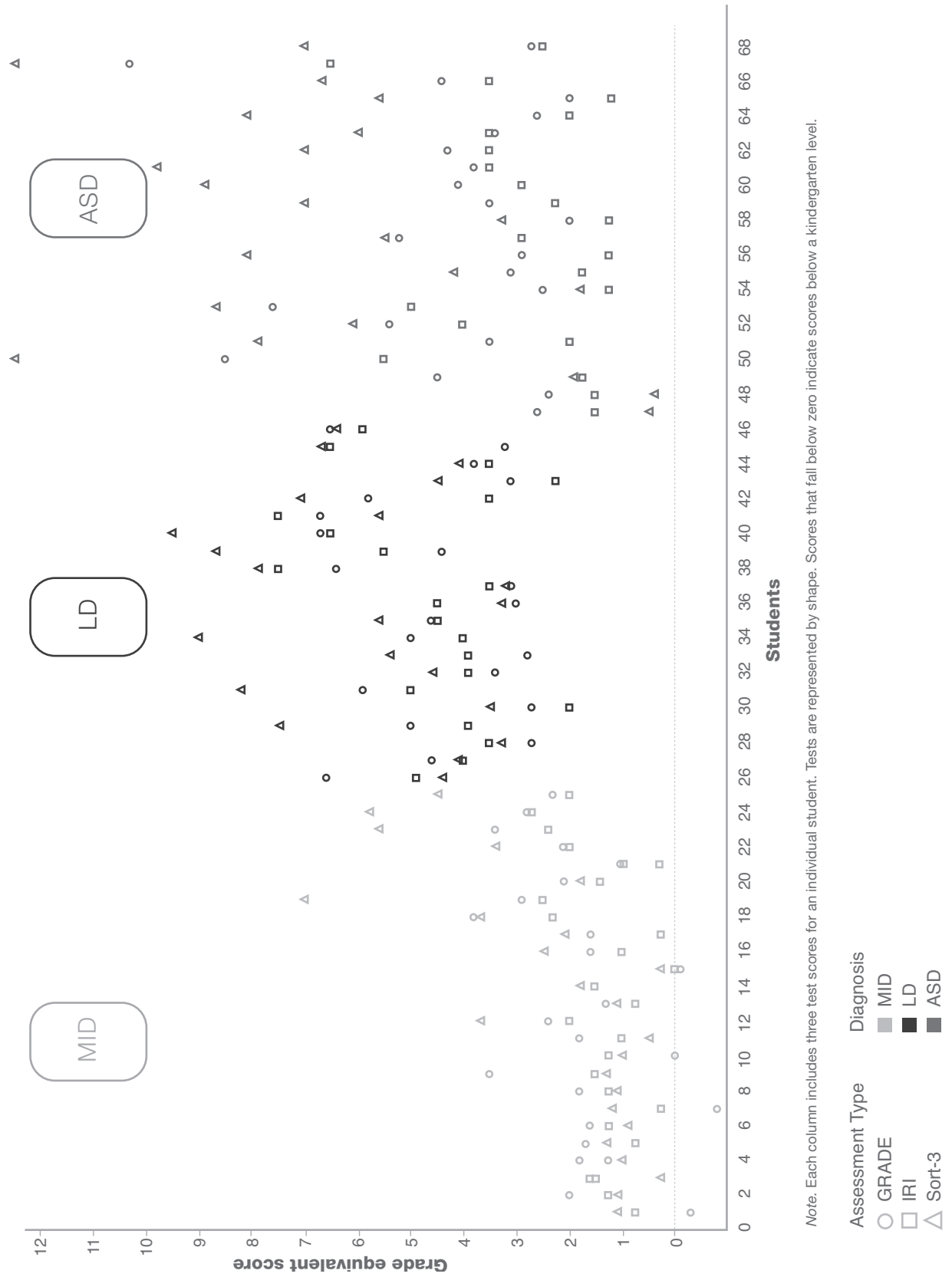
What does this mean for practitioners? Part of the draw of IRIs is that they are perceived as a more 'authentic' measure of reading in that the student is reading connected text and then answering questions in an open ended format as opposed to reading sentences and short passages alone followed by multiple choice options. However, since IRIs appear to be adding a difficulty that has little to do with reading comprehension skills, these are less 'authentic' for autistic students as they are both a reading comprehension assessment and a 'talking about what I comprehended' assessment, which is a completely different skill.

The authors agree that IRIs serve an important diagnostic function for teachers - listening to a student read and asking questions about what has been read is an excellent window into their reading ability. Additionally, being able to verbally answer questions about what one has just read is an important skill. However, it is a different and separate skill from reading comprehension. So, any reading assessment where complex oral responses are needed should be interpreted with caution, particularly in relation to autistic students. Since the gold standard psycho-educational assessment is often regarded as the individually administered assessment that typically involves social interaction, there may be an underestimation of the skills and understandings of autistic students based on the nature of the assessment. So, caution is needed in interpreting assessment results in reading and other areas of development and functioning.

References

- Arciuli J, Stevens K, Trembath D and Simpson I C (2013) The relationship between parent report of adaptive behaviour and direct assessment of reading ability in children with autism spectrum disorder *Journal of Hearing, Language and Speech Research*, 56 (6), 1837–44.
- Atkinson L, Slade L, Powell D and Levy J P (2017) Theory of mind in emerging reading comprehension: a longitudinal study of early indirect and direct effects *Journal of Experimental Child Psychology* 16, 225–238.
- Bonifacci, P and Snowling, M J (2008) Speed of processing and reading disability: a cross-linguistic investigation of dyslexia and borderline intellectual functioning *Cognition* 107(3), 999–1017.
- Braun, H I and Mislevy, R (2005) Intuitive test theory *Phi Delta Kappan* 86 (7), 488.
- Brown, J I, Fishco, V V and Hanna, G S (1993) *Nelson–Denny Reading Test* Itasca, IL: Riverside.
- Catts, H, Adlof, S and Weismer, S (2006) Language deficits in poor comprehenders: a case for the simple view of reading *Journal of Speech, Language and Hearing Research* 49(2), 278–293.
- Crocker, L and Algina, J (2008) *Introduction to classical and modern test theory* Mason, Ohio: Cengage Learning.
- Ehri, L C (2014) Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning, *Scientific Studies of Reading* 18 (1), 5–21.
- Fountas, I C and Pinnell, G S (2011) *Fountas and Pinnell benchmark assessment systems 1 and 2 (2nd ed)* Portsmouth, NH: Heinemann.
- Gough, P and Hillinger, M (1980) Learning to read: an unnatural act *Bulletin of the Orton Society* 30 (1980), 179–196.
- Gough, P and Tunmer, W (1986) Decoding, reading, and reading disability *Remedial and Special Education* 7 (1), 6–10.
- Morris, D, Bloodgood, J W, Perney, J, Frye, E M, Kucan, L, Trathen, W, Ward, D and Schiagal, R (2011) Validating craft knowledge *The Elementary School Journal* 112, 2, 205–33.
- Nation, K and Cocksey, J (2009) The relationship between knowing a word and reading it aloud in children’s word reading *Journal of Experimental Child Psychology* 103(3) 296–308.
- Paynter, J M (2015) Assessment of school-aged children with autism spectrum disorder *Journal of Psychologists and Counsellors in Schools* 25, 1, 104–115.
- Provost, M, Lambert, M and Babkie, A (2010) Informal reading inventories: creating teacher-designed literature based assessments *Intervention in School and Clinic* 45, 4, 211–220.
- Ricketts, J, Jones, C R, Happe, F and Charman, T (2012) Reading comprehension in autism spectrum disorders: the role of oral language and social functioning *Journal of Autism and Developmental Disorders* 43, 4, 807–816.
- Rindskopf, D (1996) Partitioning chi-square: something old, something new, something borrowed, but nothing blue, in A von Eye, and C Clogg (eds) *Categorical Variables in Developmental Research* New York: Academic Press.
- Saldafia, D, Carreiras, M and Frith, U (2009) Orthographic and phonological pathways in hyperlexic readers with autism spectrum disorders *Developmental Neuropsychology* 34, 3, 240–253.
- Slossen, R and Nicolson, C (2008) *SORT-R: Slossen Oral Reading Test – revised (version 3)* East Aurora, NY: Slosson Educational Publications.
- Torgensen, J (2002) The prevention of reading difficulties *Journal of School Psychology* 40, (1), 7–26.
- Tunmer, W and Chapman, J (2012) Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading* 16 (12), 122–140.
- Williams, K (2001) *Group reading assessment and diagnostic evaluation* Circle Pines, MN: American Guidance Service, Inc.
- Woodcock, RW (2011) *Woodcock reading mastery tests (third edition)* San Antonio, TX: Pearson Assessment.

Figure 1: Student performance on the three literacy measures indicating variance of scores



Note. Each column includes three test scores for an individual student. Tests are represented by shape. Scores that fall below zero indicate scores below a kindergarten level.

Assessment Type Diagnosis

○ GRADE ■ MID

□ IRI ■ LD

△ Sort-3 ■ ASD