Published in final edited form as:

J Commun Disord. 2008 ; 41(4): 358–371. doi:10.1016/j.jcomdis.2008.02.003.

Comparison of performance on two nonverbal intelligence tests by adolescents with and without language impairment

Carol A. Miller and

The Pennsylvania State University

Erin Gilbert

University of Pittsburgh

Children whose language development is not proceeding at the expected rate, but who have no cognitive deficit or other developmental disorder, are considered to have a specific language impairment (SLI). The word *specific* refers to the supposedly circumscribed nature of the deficits found in SLI: they do not have sensory, cognitive, neurological, or socio-emotional impairments that might account for their language problems. Leonard (1998) describes the history of research and clinical interest in this group of children. Although terms and definitions have changed over the years, one constant has been that the language performance of children with SLI is not commensurate with their nonverbal intellectual ability.

The idea of ruling out cognitive impairments in SLI has become instantiated in an informal but widely accepted standard of requiring children to earn a nonverbal IQ score no more than 1 SD below the mean (i.e., ≥ 85). Plante (1998) traced this custom back to the work of Stark and Tallal (1981), but noted that the context of the criterion in that study has been lost. Plante also noted that, although the nonverbal IQ criterion is intended to provide comparability of samples across studies, a score of 85 on one IQ test is not necessarily equivalent to an 85 on another test. In the current paper, we provide further evidence that nonverbal IQ scores are dependent on the test being used, and suggest that inter-test differences are not the same for all children.

Tests that are intended to measure nonverbal IQ use a variety of tasks and theoretical constructs, and also vary in their psychometric properties (DeThorne & Schaefer, 2004). Thus, different nonverbal IQ tests not only attempt to measure different abilities, but vary in the reliability and validity with which they do so. It is to be expected that a single individual will not receive the same score on two nonverbal IQ tests if only because of measurement error, but in addition, the underlying abilities being evaluated are likely to be different for any two tests.

Despite these issues, one purpose of applying nonverbal IQ criteria for SLI is to ensure that the scores of children with SLI and their typically developing peers are similar. Swisher and colleagues, however, have shown that even when similar scores are obtained on the same test, one cannot be certain that nonverbal abilities are truly equivalent. Swisher and Plante (1993) administered two nonverbal intelligence tests to children with and without SLI aged 4 to 5 years, and examined the covariance relationships among subtests, or groups of similar items, for each test. They found that these relationships were different for the SLI and normal language groups. In another study, Swisher, Plante, and Lowell (1994) administered three nonverbal IQ tests to children between the ages of 8 and 10 years. The results showed that scores on the tests

differed by as many as 10 standard score points for children with SLI and as many as 15 points for their age- and gender-matched typically developing peers. The authors further investigated variation in performance on different types of items (e.g., matching, reasoning by analogy, spatial rotation) and the correlations between item types and overall IQ score for the SLI and typically developing groups. Overall, correlations between any particular item type and total scores were lower for children with SLI, and the item types that were more strongly related to total score differed between the groups. Swisher et al. concluded that “even when IQ scores are identical for children with LI and children with NL, the numerical values may mask underlying group differences in skills.” (p. 238)

From a clinical standpoint, nonverbal IQ is often considered an indicator of an individual's potential to learn language, as documented and critiqued by Casby (1992). Swisher et al. (1994) discuss the weakness of this reasoning, and there is a substantial literature arguing against the practice of cognitive referencing (e.g., Cole, Coggins, & Vanderstoep, 1999; Cole, Dale, & Mills, 1990; DeThorne & Watkins, 2006; Fey, Long, & Cleave, 1994). From a research perspective, setting a nonverbal IQ criterion is, as indicated earlier, a way of ensuring the specificity of SLI. However, a cutoff of ≥ 85 appears arbitrary, given findings such as the data of Swisher and Plante (1993) and Swisher et al. (1994), the large literature demonstrating that children with SLI have weaknesses in nonlinguistic cognitive skills (see Leonard, 1998 for a review), and evidence that children above and below the cutoff do not differ in their language profiles (Tomblin & Zhang, 1999) or in their response to intervention (Cole et al., 1999; Fey et al., 1994). Arguably, all a researcher needs to do is rule out mental retardation, in which case the cutoff should be set at 70 but take into account the standard error of measurement of the test being used (APA, 2000).

The purpose of the present study is to compare scores obtained by the same children on two different nonverbal IQ instruments, and determine whether different tests lead to different diagnostic classifications. Although group differences have been investigated previously, their consequences for diagnostic classification have not. The study involves a relatively large sample, consisting of 204 young adolescents who, at 14 years of age, are older than children in previous studies comparing nonverbal IQ tests. The participants in the current study also represent a wide range of language and cognitive abilities; not only were there typically developing children and children with SLI, but also children with IQs below the SLI cut-off (the NLI group) and a small number of children with low nonverbal IQ in the presence of typical language. We asked the following research questions.

1. Are there differences in the performances of the same children on two different tests of nonverbal IQ, and if so, are the differences greater for children with language impairment?
2. Do children receive the same diagnostic classifications when two different nonverbal IQ tests are used as criteria?

Methods

Participants

The participants in this study were 204 children (mean age of 13.9 years) who were enrolled in a large-scale longitudinal study of language impairment. See Tomblin and Zhang (2006) for a full description of the sampling and testing procedures. The children were a subset of the full sample in the longitudinal study. They were selected to participate in a study of response times at grades 3 and 8 (for details, see Miller, Kail, Leonard, & Tomblin, 2001; Miller et al., 2006). The subset was selected at grade 3 by including all children who had previously met the criteria for language impairment (described below) and randomly sampling those who did not. All children from this subsample who remained in the longitudinal study at grade 8 were

included in the present study. At grade 8, they took two tests of nonverbal IQ; one was used for diagnostic purposes, the other was not. These test scores provided the data to be examined here. The second test was administered in order to obtain a nonverbal IQ measure that did not rely on timed performance, given the focus on experimental response time tasks. The present study takes advantage of the opportunity to examine performance on two tests given concurrently. The tests are described in more detail in the “Instruments” section.

In grades 2, 4, and 8, the participants, like all children in the longitudinal study, received a battery of diagnostic tests that assessed grammar, vocabulary, and discourse in the receptive and expressive modalities. They were also administered selected subtests from either the Wechsler Preschool and Primary Scale of Intelligence-Revised (Wechsler, 1989) or the Wechsler Intelligence Scale for Children-Third Edition (Wechsler, 1991). At grade 8, the Picture Completion and Block Design subtests of the WISC-III were administered. This test battery was used to classify the participants into categories. The details of the diagnostic scheme, which is summarized here, can be found in Tomblin, Records, and Zhang (1996).

A cutoff for impaired performance was established in two areas, language and nonverbal IQ. Composite language scores were derived from the receptive and expressive measures in each domain of language (grammar, vocabulary, discourse). These scores were standardized for the entire longitudinal sample of 604 children, yielding 5 Z-scores for each participant. Children who scored more than -1.25 standard deviations below the mean on two or more of the composite measures were classified as language impaired. A standard nonverbal IQ score was derived from the two subtests of the WISC-III. A score of 86 or above was considered “normal,” as a score of 85 was not possible in the computation of IQ from two subtests.

These two criteria, language and nonverbal IQ, produced a four-way classification. Participants who scored above the cutoff in both areas were considered typically developing, designated here as NLD. Those who were above the cutoff in nonverbal IQ but below it in language were considered SLI, and those who scored below the cutoffs in both areas were classified as non-specifically language impaired, or NLI. A small number of children were below the cutoff for nonverbal IQ but above the cutoff for language; this low cognitive group is designated COG. Of the 204 participants in the current study, 116 were NLD, 51 SLI, 27 NLI, and 10 COG according to testing at grade 8.

Instruments

As described above, the Picture Completion and Block Design subtests of the WISC-III were used for the purposes of diagnostic classification. Short forms of the Wechsler scales are often used for research purposes (Kaufman, Kaufman, Balgopal, & McLean, 1996; Sattler, 2001). Kaufman et al. (1996) reported that the Picture Completion-Block Design dyad was a satisfactory short form for Performance IQ, with a reliability coefficient of 0.88 and correlation with Performance IQ of 0.86 for 14-year-olds.

Verbal directions are used in both the Picture Completion and Block Design subtests. Picture Completion uses a set of colored pictures, each of which is missing a part which the child is asked to identify. This subtest is designed to test fluid intelligence through tapping into perceptual organizational skills (Wechsler, 1991). The second performance subtest, Block Design, uses a set of two-dimensional geometric patterns, either modeled by the examiner or printed in the stimulus book, which the child must replicate as closely as possible. This subtest is also designed to test perceptual organizational skills (Wechsler, 1991). Both subtests involve time limits for responding, and in Block Design, the examinee receives a higher score for a rapid, correct response.

The Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) was also administered. This test was developed to meet the needs of measuring general intelligence solely through nonverbal means; item response and presentation are completely nonverbal. The Symbolic Memory and Analogic Reasoning subtests were used. They are intended to measure, respectively, memory and reasoning for symbolic content. Neither subtest places time limits on responses. The UNIT includes an Object Memory subtest that also measures memory for symbolic content, but its reliability is lower than for Symbolic Memory (Bracken & McCallum, 1998). For 14-year-olds, correlations of the subtests with the Standard Battery Full Scale IQ are 0.73 for Symbolic Memory and 0.77 for Analogic Reasoning.

The Symbolic Memory subtest uses universal black or green symbols of a baby, girl, boy, woman, and man. An examinee is shown a sequence of these symbols for 5 seconds and then required to reconstruct the previously viewed sequence. The purpose of the Symbolic Memory Subtest is to measure “short-term visual memory and complex sequential memory for meaningful material” (Bracken & McCallum, 1998, p. 3). The Analogic Reasoning subtest consists of presenting “incomplete conceptual or geometric analogies in a matrix format” (p. 3) that involve common objects and novel geometric figures. The examinee is asked to respond by pointing to the correct choice from four options to complete the analogy. The purpose for this subtest is to measure symbolic reasoning as a way to assess fluid intelligence (Bracken & McCallum, 1998). Neither of the subtests is timed, but the stimulus sequence in Symbolic Memory is exposed for 5 seconds and then covered.

The UNIT is less well-known in the language disorders literature than the Wechsler scales. However, in a recent review of 16 tests of nonverbal ability, DeThorne and Schaefer (2004) included both the UNIT and the WISC-IV, successor to the WISC-III, among the four tests recommended for use in evaluating children with language impairment, based on overall design and psychometric properties. Table 1 provides a summary of several reliability characteristics of the subtests used in the current study, as well as composite scores. Each test manual also provides several measures of validity. Most of these measures are not reported for individual subtests, or for the combination of two subtests used in the current study. Selected data are presented here to confirm that the tests have generally acceptable validity.

In concurrent validity studies (Wechsler, 1991), the WISC-III correlated moderately to highly with two other Wechsler tests, the WISC-R (Wechsler, 1974) and Wechsler Adult Intelligence Scale-Revised (WAIS-R, Wechsler, 1981). Table 2 shows the correlations with these tests for the WISC-III performance IQ (PIQ) and the two relevant subtests, as well as the correlation between WISC-III PIQ and the Nonverbal Reasoning Scale of the Differential Ability Scales (Elliott, 1990). All correlations were moderate to high.

The UNIT manual (Bracken & McCallum, 1998) reported concurrent validity with the WISC-III for a sample of children with learning disabilities. The corrected correlation between the UNIT Standard Battery Full Scale score and the WISC-III PIQ was 0.76. In addition, correlations with the Woodcock-Johnson Tests of Achievement –Revised (WJ-R, Woodcock & Johnson, 1989/1990) were compared for the UNIT and WISC-III in a sample of children with learning disabilities. The UNIT Standard Battery Full Scale score had corrected correlations of 0.51 and 0.37 with the WJ-R Broad Knowledge and Skills composites, respectively. The WISC-III PIQ correlated with Broad Knowledge at 0.67 and Skills at 0.18. The concurrent validity of the UNIT was also evaluated using the Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1990). The corrected correlation between the UNIT Standard Battery Full Scale score and the K-BIT IQ Composite was 0.82.

Both the WISC-III and UNIT reported test results from samples of children with speech-language disorders. For the WISC-III, only composite scores were reported. A sample of “44

multiply handicapped children aged 6-13 years (median = 9 years) with primary diagnoses of expressive or receptive language disorders” (Wechsler, 1991, p. 215) had a mean PIQ of 78.2 ($SD = 17.1$) and a mean verbal IQ of 69.2 ($SD = 12.4$). For the UNIT, “a sample of 57 examinees diagnosed with communication disorders and who were receiving special services for persons with speech and language impairment” (Bracken & McCallum, 1998, p. 163) were compared to a matched control group. For Symbolic Memory, the clinical group's mean was 9.49 ($SD = 3.04$) and the control group's mean was 10.04 ($SD = 3.12$). For Analogic Reasoning, the mean of the clinical group was 8.40 ($SD = 3.35$) and the mean of the control group was 10.04 ($SD = 3.06$). In summary, both the WISC-III and UNIT appear to have adequate reliability and validity, and to be suitable for testing individuals with language impairment.

Procedure

The tests, as part of a larger battery of tests and tasks, were administered individually by trained examiners who visited the child's school or home in a customized van. The WISC-III subtests were administered before the UNIT subtests, and usually on different days, although in a few cases, more than one session was conducted on the same day for logistical reasons.

Results

Table 3 shows means and standard deviations for IQ scores and subtest scores in each group. It is clear from inspection of the means that UNIT IQ scores are noticeably higher for the NLI and COG groups. A mixed-model ANOVA with diagnostic category as the between-subjects factor and test score as the within-subjects factor confirmed this observation. There was a significant main effect of group ($F(3, 200) = 53.06, p < .001$, partial $\eta^2 = 0.44$), and a significant main effect of test ($F(1, 200) = 12.38, p < .001$, partial $\eta^2 = 0.06$). The group by test interaction was significant ($F(3, 200) = 15.88, p < .001$, partial $\eta^2 = 0.19$), indicating that the difference between test scores was not the same for all groups. Post-hoc analyses using the unequal N HSD test showed that the NLD group's scores on the WISC-III and UNIT did not differ at $\alpha = .05$, but the SLI scored higher on the WISC-III, and the NLI and COG groups scored higher on the UNIT.

Difference scores were computed by subtracting the UNIT IQ score from the WISC-III score for each individual. Boxplots showing the distribution of difference scores for each diagnostic category (with WISC-III score as criterion) are found in Figure 1. The median differences were 3 for NLD, 2 for SLI, -14 for NLI, and -15.5 for COG.

To further explore relationships between the tests for the four groups, a series of Pearson correlations were computed. Correlations of WISC-III scores with UNIT scores are shown in Table 4. For the entire group, $r = 0.41, p < .001$. This is a smaller correlation than the 0.76 reported by Bracken and McCallum (1998) between the UNIT Full Scale IQ and the WISC-III Performance IQ for children with learning disabilities. Scores on the two tests were significantly correlated only for the NLD group, and this correlation was of medium magnitude (Cohen, 1988). The COG group had a correlation of similar magnitude, but due to the small n , it did not reach significance, and should be interpreted with caution. The association between scores was more noticeably attenuated in the SLI and NLI groups.

In order to determine how the differences in IQ scores could affect classification of individuals, the numbers of individuals in each category according to each test were cross-tabulated, as shown in Table 5. The top panel of Table 5 shows the results of applying a cut-off of ≥ 85 for each test (for the WISC-III, 86 was the cut-off, but was functionally the same as 85, because the latter score could not occur). A small number of individuals in the NLD group, only 6%, would have been classified as COG using the UNIT, and 16% of the SLI group were reclassified

as NLI. In contrast, 74% of the NLI group were reclassified as SLI, and 80% of the COG group were reclassified as NLD.

It is possible that with different cut-offs, fewer individuals would be classified differently by the two tests. The results of setting cut-offs at 80 and at 70 are shown in the middle and bottom panels, respectively, of Table 5. As the cut-off goes down, more children are included in the NLD and SLI categories, and fewer in NLI and COG, by definition. With 80 as the cut-off value, again relatively small percentages of individuals with NLD (2%) and SLI (12%) changed classification, compared to 72% of the NLI group. Only 3 individuals were classified as COG by the WISC-II at this cut-off and 2 of them were reclassified by the UNIT. When 70 was used as the cut-off, 4% of the NLD group and 11% of the SLI group changed. No one was classified as COG by the WISC-III at this cut-off, and only 3 were classified as NLI; all 3 were reclassified as SLI according to the UNIT.

Discussion

The first research question asked whether there were differences in the performances of the same children on two different tests of nonverbal IQ, and if so, whether the differences were greater for children with language impairment. The results suggested that performance on the two tests was different for all except the typically developing group, which was unique in showing no significant difference between mean test scores as well as a significant correlation between the two tests. The most extreme discrepancies between test scores were found in the NLI and COG groups. The SLI group demonstrated a significant mean difference between their scores on the two tests, but the distribution of their difference scores was more similar to the NLD group than to the NLI and COG groups (see Figure 1). The average discrepancies between the test scores were not tightly linked to language status, as the COG group (with normal language) was more similar to the NLI group and the SLI group (with low language) was more similar to the NLD group.

Some differences between test scores are to be expected due to measurement error alone. The test-retest reliability of the subtests is reasonable, but far from perfect; therefore, we would expect discrepancies between two administrations of the same test, as well as between two different tests. Also, because the children were initially categorized based on their WISC-III scores, regression to the mean might contribute to the higher UNIT scores of the lower-scoring NLI and COG groups. Still, the differences between the tests are unlikely to be due entirely to measurement error. Standard errors of measurement are not available for the pairs of subtests that were used, but based on the SE_M s for the subtests and for composite scores, they probably would not be much greater than 5. The mean and median group differences for the NLI and COG groups were more than twice that value.

The source of differences between the two tests, whether due to error, true score differences, or both, is to some extent irrelevant. The differences exist, and the examination of the second research question showed that these differences led to different diagnostic classifications for a non-trivial number of individuals. Few children with NLD changed classification when the UNIT was used as a criterion in place of the WISC-III, and this result is not surprising given the small differences between test scores for the NLD group. Depending on the cut-off used, 11 to 16% of individuals in the SLI group were reclassified as NLI. More strikingly, nearly three-quarters of the children with NLI were reclassified as SLI, and most of the children with COG status were reclassified as NLD. These results suggest that comparability of participants across studies cannot be assumed when different tests are used. The current findings add two more tests to the list of those that have been shown not to yield equivalent scores (Swisher & Plante, 1993; Swisher et al., 1994). Far from trying to construct lists of “good” or “bad” tests, however, our interpretation is a more general one: any two tests are likely to yield different

results for a sample of children who are not typically developing. It follows, therefore, that children with SLI in one study are not necessarily comparable to children with SLI in another study where a different nonverbal IQ test was used.

The results of the current analyses do not tell us why scores differed on the two tests. The tests themselves differ in several ways. The Block Design and Picture Completion subtests from the WISC-III are both intended to assess perceptual organizational skills (Wechsler, 1991). The subtests are presented to the examinee with a set of verbal directions and an example given by the examiner. For the Block Design subtest, bonus points are awarded for correct answers given within a certain time limit. It is possible that the examinee may use verbal mediation on the Picture Completion subtest but it seems unlikely that verbal mediation would be used for the Block Design subtest. For both of the WISC-III subtests time constraints are emphasized and verbal instructions are used.

According to the test developers, the Symbolic Memory and Analogic Reasoning subtests of the UNIT assess three different skills (Bracken & McCallum, 1998). The Symbolic Memory portion of the UNIT evaluates short-term visual memory and complex sequential memory and the Analogic Reasoning subtest taps into the examinee's symbolic reasoning skills. Both of the subtests have demonstration and sample items at the beginning of the subtest to clarify the examinee's understanding. There are also checkpoint items to provide additional opportunities to ensure that the participant understands the task. No bonus points are awarded and there are no time constraints given for the examinee to provide an answer, although in Symbolic Memory, the stimuli are exposed for a limited time. Non-verbal directions are given to the participant through the use of gestures and the stimulus book. It appears to be possible that the examinee may use verbal mediation in both the Symbolic Memory and Analogical Reasoning subtests.

The WISC-III and UNIT subtests described here, then, differ in potentially important ways. The underlying abilities that they are intended to measure are different, but they also differ in more superficial (but more observable) aspects. Speedy responses are rewarded on the WISC-III but not the UNIT (note, however, that the complete nonverbal batteries of each test include both timed and untimed subtests). The UNIT is completely nonverbal, in that the examiner's instructions as well as the examinee's responses are given nonverbally, whereas the WISC-III employs verbal instructions.

Research is needed to determine how such differences between tests affect the performance of clinical groups and individuals. Intuitively, it seems that the use of verbal directions could place children with language impairment at a disadvantage. In this study, however, children with SLI performed slightly better on the WISC-III, making a simple explanation relating verbal instructions to poorer performance unlikely. Time constraints on tests might also affect performance differentially. Children with SLI have been found to have slower processing speed than typically-developing peers on many tasks, including some that do not involve language (Kail, 1994; Kohnert & Windsor, 2004; Miller et al., 2001; Miller et al., 2006; Montgomery, 2005; Windsor & Hwang, 1999). Miller and colleagues (2001; 2006) found that children with NLI were also slow processors. If speed of processing is a problem for children with language impairment, they may be at a greater disadvantage on tests that reward rapid responding. Again, however, the children with SLI in this study, as a whole, were not adversely affected by the time limits of the WISC-III.

The finding that children with SLI performed better on the WISC-III whereas children with NLI and COG performed better on the UNIT suggests that the interaction between characteristics of tests and characteristics of individuals is complex. This interpretation is consistent with research by Colozzo and Johnston (2004) suggesting that use of a verbal

strategy for a supposedly nonverbal task hinders performance for most, but not all, children with SLI. Thus, the degree to which a particular task elicits or favors verbal mediation is a factor that may be important in understanding differences in nonverbal IQ scores, but is more difficult to evaluate than the use of verbal instructions or the presence of time limits.

The foregoing discussion is not intended to suggest that one test is better than another. Rather, we conclude, based on our results and the literature on this topic, that there is no one test that is a “better” measure of nonverbal intelligence for all individuals. The assumption that nonverbal intelligence is a construct that can be reliably assessed with different instruments may hold for typically developing children (although even that is debatable), but the assumption does not hold for children with language impairment, particularly those who score relatively low on one or more nonverbal IQ tests. This study complements a growing body of evidence that nonverbal IQ is not a good predictor of who will benefit from language intervention (Cole et al., 1999; Cole et al., 1990; DeThorne & Watkins, 2006; Fey et al., 1994; Swisher & Plante, 1993; Swisher et al., 1994). Therefore, it behooves clinicians to be prepared to discuss issues regarding nonverbal IQ and its use as a criterion for eligibility for services. It is also incumbent upon researchers studying children with SLI to carefully consider and justify decisions about whether or not to assess nonverbal IQ, and the purpose of such testing (e.g., to rule out mental retardation, to equate groups). If nonverbal IQ is assessed, the choice of test should be informed by a good understanding of the available options (DeThorne & Schaefer, 2004) as well as the limitations of such testing.

References

- APA. Diagnostic and Statistical Manual of Mental Disorders. Washington D.C.: Author; 2000.
- Bracken, B.; McCallum, S. Universal Nonverbal Intelligence Test. Itasca, IL: Riverside; 1998.
- Casby MW. The cognitive hypothesis and its influence on speech-language services in schools. *Language, Speech and Hearing Services in Schools* 1992;23(3):198–202.
- Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- Cole KN, Coggins TE, Vanderstoep C. The influence of language/cognitive profile on discourse intervention outcome. *Language Speech and Hearing Services in Schools* 1999;30(1):61–67.
- Cole KN, Dale PS, Mills PE. Defining language delay in young children by cognitive referencing - Are we saying more than we know. *Applied Psycholinguistics* 1990;11(3):291–302.
- Colozzo, P.; Johnston, J. Working memory: The capacity-strategy connection. Paper presented at the Symposium on Research in Child Language Disorders; Madison, WI. Jun. 2004
- DeThorne LS, Schaefer BA. A guide to child nonverbal IQ measures. *American Journal of Speech-Language Pathology* 2004;13:275–290. [PubMed: 15719895]
- DeThorne LS, Watkins RV. Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics & Phonetics* 2006;20(9):641–658. [PubMed: 17342874]
- Elliott, CD. Differential Ability Scales: Administration and scoring manual. San Antonio, TX: Psychological Corporation; 1990.
- Fey, ME.; Long, SH.; Cleave, P. Reconsideration of IQ criteria in the definition of specific language impairment. In: Watkins, RV.; Rice, M., editors. *Specific language impairments in children*. Vol. 4. Baltimore, MD: Brookes; 1994. p. 161-178.
- Kail R. A method of studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech and Hearing Research* 1994;37:418–421. [PubMed: 8028323]
- Kaufman AS, Kaufman JC, Balgopal R, McLean JE. Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology* 1996;25(1):97–105.
- Kaufman, AS.; Kaufman, NL. Kaufman Brief Intelligence Test manual. Circle Pines, MN: American Guidance Service; 1990.

- Kohnert K, Windsor J. The search for common ground: Part II Nonlinguistic performance by linguistically diverse learners. *Journal of Speech, Language, and Hearing Research* 2004;47:891–903.
- Leonard, LB. *Children with specific language impairment*. Cambridge, MA: MIT Press; 1998.
- Miller CA, Kail R, Leonard LB, Tomblin JB. Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 2001;44:416–433.
- Miller CA, Leonard LB, Kail RV, Zhang X, Tomblin JB, Francis DJ. Response time in 14-year-olds with language impairment. *Journal of Speech, Language, and Hearing Research* 2006;49(4):712–748.
- Montgomery J. Effects of input rate and age on the real-time language processing of children with specific language impairment. *International Journal of Language & Communication Disorders* 2005;40(2): 171–188. [PubMed: 16101273]
- Plante E. Criteria for SLI: The Stark and Tallal legacy and beyond. *Journal of Speech, Language, and Hearing Research* 1998;41(4):951–957.
- Sattler, JM. *Assessment of children: Cognitive applications*. Vol. Fourth. San Diego: Jerome M. Sattler; 2001.
- Stark RE, Tallal P. Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders* 1981;46:114–122. [PubMed: 7253588]
- Swisher L, Plante E. Nonverbal IQ tests reflect different relations among skills for specifically language-impaired and normal children. *Journal of Communication Disorders* 1993;26:65–71. [PubMed: 8340491]
- Swisher L, Plante E, Lowell S. Nonlinguistic deficits of children with language disorders complicate the interpretation of their nonverbal IQ scores. *Language, Speech and Hearing Services in Schools* 1994;25:235–240.
- Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research* 1996;39:1284–1294. [PubMed: 8959613]
- Tomblin, JB.; Zhang, X. Language patterns and etiology in children with specific language impairment. In: Tager-Flusberg, H., editor. *Neurodevelopmental disorders*. Cambridge, MA: MIT Press; 1999.
- Tomblin JB, Zhang X. The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research* 2006;49(6):1193–1208.
- Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: Psychological Corporation; 1974.
- Wechsler, D. *Manual for Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation; 1981.
- Wechsler, D. *WPPSI-R Manual: Wechsler Preschool and Primary Scale of Intelligence-Revised*. New York: Psychological Corporation; 1989.
- Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation; 1991.
- Windsor J, Hwang M. Testing the generalized slowing hypothesis in specific language impairment. *Journal of Speech, Language, and Hearing Research* 1999;42:1205–1218.
- Woodcock, RW.; Johnson, MB. *Woodcock-Johnson Psycho-Educational Battery-Revised*. Itasca, IL: Riverside Publishing; 1989/1990.

Figure 1. Distributions of differences between WISC-III and UNIT scores by group. Boxplots show medians (square marker), 25th to 75th percentile range (box), non-outlier range (whiskers), and outliers (round marker).

Table 1
Reliability data for WISC-III and UNIT subtests for age 14 years

Reliability measures	WISC-III			UNIT		
	Picture Completion	Block Design	Performance IQ	Symbolic Memory	Analogic Reasoning	Full Scale IQ (Standard Battery)
Internal consistency (split-half) ^a	0.72	0.90	0.89	0.87	0.78	0.94
Test-retest stability ^b	0.82	0.83	0.88	0.79	0.73	0.91
Standard error of measurement ^a	1.59	0.95	4.97	1.07	1.29	3.68

Note. Data from test manuals (Bracken & McCallum, 1998; Wechsler, 1991).

^aComputed for age 14.

^bComputed for ages 14-15 for WISC-III and ages 14-17 for UNIT

Table 2
Correlations (corrected) between WISC-III and WISC-R, WAIS-R, and DAS

	WISC-R	WAIS-R	DAS Nonverbal Reasoning
PIQ	0.81	0.80	0.78
Picture Completion	0.57	0.46	--
Block Design	0.76	0.79	--

Note. WISC-R: Wechsler Intelligence Scale for Children-Revised; WAIS-R: Wechsler Adult Intelligence Scale-Revised; DAS: Differential Ability Scales.

Table 3
Means (SDs) for IQ scores and subtest scores by group

Group	WISC-III IQ	WISC-III Block Design	WISC-III Picture Completion	UNIT IQ	UNIT Symbolic Memory	UNIT Analogic Reasoning
NLD	101.6 (9.7)	10.6 (2.2)	9.9 (2.3)	99.9 (9.9)	10.6 (1.2)	9.4 (2.4)
SLI	97.9 (10.3)	10.2 (2.4)	9.1 (2.1)	92.8 (10.4)	9.2 (2.1)	8.4 (2.4)
NLI	75.8 (5.8)	5.3 (2.6)	6.3 (2.1)	87.3 (10.0)	9.4 (1.9)	6.6 (2.8)
COG	80.0 (3.7)	5.9 (2.1)	7.1 (1.5)	92.4 (11.2)	9.1 (1.5)	8.4 (2.4)

Table 4
Correlations between WISC-III score and UNIT score

Group	Pearson <i>r</i>
SLI	0.11
NLI	0.20
NLD	0.31*
COG	0.30
All groups	0.41**

*
 $p < .01$.

**
 $p < .001$.

Table 5
Comparison of diagnostic classifications using WISC-III and UNIT scores

Classification using WISC-III		Classification using UNIT		
	SLI	NLI	NLD	COG
		Cut-off: 85		
SLI	43	8		
NLI	20	7		
NLD			109	7
COG			8	2
		Cut-off: 80		
SLI	53	7		
NLI	13	5		
NLD			120	3
COG			2	1
		Cut-off: 70		
SLI	67	8		
NLI	3	0		
NLD			122	4
COG			0	0

Note. Blank cells are not possible combinations, as language status does not change.