

Spring 2015

# Evidence-Based Practices in Selecting Standardized Assessments of Child Language

Darielle Cooper  
*James Madison University*

Follow this and additional works at: <http://commons.lib.jmu.edu/honors201019>



Part of the [Speech Pathology and Audiology Commons](#)

---

## Recommended Citation

Cooper, Darielle, "Evidence-Based Practices in Selecting Standardized Assessments of Child Language" (2015). *Senior Honors Projects, 2010-current*. 23.

<http://commons.lib.jmu.edu/honors201019/23>

This Dissertation/Thesis is brought to you for free and open access by the Honors College at JMU Scholarly Commons. It has been accepted for inclusion in Senior Honors Projects, 2010-current by an authorized administrator of JMU Scholarly Commons. For more information, please contact [dc\\_admin@jmu.edu](mailto:dc_admin@jmu.edu).

Evidence-Based Practices in Selecting Standardized Assessments of Child Language

---

An Honors Program Project Presented to  
the Faculty of the Undergraduate  
College of Health and Behavioral Studies  
James Madison University

---

by Darielle Cooper

May 2015

---

---

Accepted by the faculty of the Department of Communication Sciences and Disorders, James Madison University, in partial fulfillment of the requirements for the Honors Program.

FACULTY COMMITTEE:

HONORS PROGRAM APPROVAL:

---

Project Advisor: Stacey Pavelko, Ph.D.,  
Director of Clinical Education, Communication  
Sciences and Disorders

---

Philip Frana, Ph.D.,  
Interim Director, Honors Program

---

Reader: Vicki Reed, Ed.D.,  
Professor, Communication Sciences and Disorders

---

Reader: Carol Dudding, Ph.D.,  
Director of SLP Graduate Program, Communication  
Sciences and Disorders

---

---

PUBLIC PRESENTATION

This work was accepted for presentation, in part or in full, at the Speech and Hearing Association of Virginia on March 27, 2015.

This work is accepted for presentation, in part or in full, at the JMU Honors Symposium on April 24, 2015.

## Table of Contents

Acknowledgments	3
Abstract	4
Introduction	5
Methods	9
Findings	12
Discussion and Clinical Implications	18
References	23

## **Acknowledgments**

I would like to express sincere gratitude toward my professor, honors thesis advisor, and soon to be colleague, Dr. Stacey Pavelko, for making this undergraduate research experience possible. I am grateful for her recognition of my potential as an honors student and for the mentorship and commitment she allocated me as a student.

I would also like to express the deepest appreciation to the faculty and professors within the Communication Sciences and Disorders department at James Madison University who taught, supported, and prepared me for the field of speech-language pathology. A special thank you goes to my committee members, Dr. Vicki Reed and Dr. Carol Dudding for volunteering as readers of this thesis. You all have profoundly impacted my academic and career aspirations.

I wish to acknowledge the helpful contribution of the James Madison University librarian, Stefanie Warlick for her assistance during the literature searches conducted in this study. Thank you to those individuals that provided us with copies of the EOWPVT and ROWPVT for use in this study.

Last but not least, thank you to my wonderful friends, fellow honors students and colleagues, Katie Gallaher, Dakota Sharp, and Megan Taliaferro, for the unconditional love and support you have given me these past four years.

## **Abstract**

From the results of a survey given to 1,399 speech-language pathologists, 11 frequently used language assessment tools for preschool and school age children were selected to undergo a systematic review. The examiner's manual of each assessment was examined to determine whether the test reported adequate psychometric properties and acceptable levels of diagnostic accuracy. Results indicated that all reviewed assessments met at least 60% of McCauley and Swisher's (1984) psychometric criteria for evaluating norm-referenced tests. Six of the 11 assessments provided information on diagnostic accuracy in the examiner's manual or in the literature. Four tests published acceptable levels of diagnostic accuracy, based on the recommendations of Plante and Vance (1994). Findings of the current study reveal that many child language assessment tools either have unknown or unacceptable levels of diagnostic accuracy. Clinical implications and further suggestions are discussed.

## **Introduction**

Speech-language pathologists (SLPs) serving preschool and school-age children are trusted to use their clinical judgment to make well-informed and ethically responsible decisions in the diagnostic component of clinical practice. Because the results of a diagnostic evaluation may determine a child's eligibility for special education services, the diagnostic process deserves a great deal of attention (Plante & Vance, 1994; Friberg, 2010).

In 2010, the American Speech-Language Hearing Association (ASHA) adopted updated roles and responsibilities for SLPs in schools. One update focuses on evolving professional practice and includes evidence-based practices (EBP) as a major tenant. The EBP framework originated in clinical medicine, but has been adopted by a variety of professions to better defend and support decisions within those respective fields. In the field of speech-language pathology, EBP incorporates the use of the best current evidence, clinical expertise, and client values to provide high-quality services. When applied to clinical assessment, Dollaghan (2004) encourages clinicians to examine both the psychometric properties and diagnostic accuracy of standardized tests to determine whether those tests meet standards of scientific rigor.

Many SLPs use standardized assessments as the objective measure in combination with subjective measures, such as, patient/caregiver survey, and clinical judgment (Plante & Vance, 1994) as part of a comprehensive assessment battery. The objective and standardized evaluation of norm-referenced assessments assures test users that scores received on a given assessment accurately reflect skill level, rather than extraneous factors (McCauley & Swisher, 1984). In general, clinicians consider these assessments appropriate and even necessary to identify language impairment; however, clinicians must use and interpret the results of standardized assessments carefully. The widespread use of these tools assumes they correctly measure the

presence or absence of impairment, although this assumption does not always hold true (Friberg, 2010). Thus, the need for clinicians to understand issues related to a standardized test's psychometric properties and diagnostic accuracy is essential to evidence based clinical practice.

### **Psychometric Properties**

An assessment's psychometric properties, such as validity and reliability, refer to those aspects of a test that state how appropriate it is for a particular purpose. Test validity, the degree to which a test accurately measures what it is intended to measure, requires data to support its purpose (Plante & Vance, 1994). In order to review evidence of validity and other psychometric properties in standardized language assessments, McCauley and Swisher (1984) developed ten criteria and evaluated 30 language and articulation tests for preschool children. McCauley and Swisher's (1984) results indicated that only two out of the 30 tests met five of the criteria employed – the first indication that standardized tests might not be as valid as test users assume.

Research immediately following McCauley and Swisher's (1984) study revealed that a variety of standardized language assessments still did not meet many of the recommended psychometric criteria (Plante & Vance, 1994). For example, Plante and Vance (1994) reviewed 21 tests of language skills for psychometric criteria and reported that only 38% of the tests met five or more criteria. As a second part of the study, Plante and Vance administered four tests that met a relatively large number of criteria to children with specific language impairment and age-matched controls. Results indicated that only one of the four tests provided acceptable levels of diagnostic accuracy (Plante & Vance, 1994). Therefore, research suggests that even tests that meet a relatively high number of psychometric criteria are not sufficient for accurate identification of language impairment because the test may not discriminate between normal and impaired language in children (Plante & Vance, 1994; Gray, Plante, Vance, & Henrichsen, 1999).

## **Diagnostic Accuracy**

In addition to considerations of psychometric properties, several authors have advocated that diagnostic accuracy must also be considered when choosing a standardized test. Diagnostic accuracy refers to the precision and ability of an assessment to accurately identify the presence or absence of a disorder. Two different measures capture diagnostic accuracy: sensitivity and specificity. When applied to speech-language pathology, *sensitivity* is the ability of the test to accurately identify individuals with disordered speech and/or language skills; whereas, *specificity* is the ability of the test to accurately identify typical speech/language development (Dollaghan, 2004; Spaulding et al., 2006). Sensitivity and specificity values range from 0 to 1.0, with 1 indicating 100% accurate identification (Friberg, 2010). These values need to be high enough to indicate that they can accurately discriminate between normal and impaired language (Gray, Plante, Vance, & Henrichsen, 1999). This aspect of the diagnostic evaluation is important for clinicians to consider because the misidentification of impairment can have drastic implications on whether the child receives appropriate intervention.

Plante and Vance (1994) suggest that 90% accuracy should be considered good, 80-89% accuracy should be considered fair, and tests with accuracy below 80% should not be used because “misidentifications occur at unacceptably high rates” (p. 21). Unfortunately, findings in the literature demonstrate that many assessments do not have acceptable levels of diagnostic accuracy (Plante & Vance, 1994; Gray, Plante, Vance, & Henrichsen, 1999; Friberg, 2010). Data also suggests that many clinicians do not consider psychometric properties or diagnostic accuracy when selecting standardized tests (Betz, Eickhoff, & Sullivan, 2013).



## **Current Research and Purpose of Study**

Betz, Eickhoff, and Sullivan (2013) investigated the factors that drive speech-language test selection in school-based clinical practice. The results of this study indicated that psychometric criteria did not appear to influence how often a test is selected. Specifically, the researchers found a lack of correlation between frequency of test use and test accuracy. Publication year was the only test characteristic that correlated significantly with the frequency of test use. Because this study focused on standardized tests used to diagnose specific language impairment (SLI), the authors did not expect these results, as a test's ability to make an accurate diagnosis would be particularly pertinent for this purpose (Betz, Eickhoff, & Sullivan, 2013).

In addition to how little influence psychometric criteria has on test selection, Betz, Eickhoff, and Sullivan (2013) found the most frequently used standardized tests were omnibus and single-word vocabulary measures. This is particularly concerning because previous literature has documented the poor diagnostic accuracy of single-word vocabulary tests in predicting status of SLI (Gray, Plante, Vance, & Henrichsen, 1999). Similarly, a more recent survey indicated that vocabulary measures are used by more than 25% of clinicians to identify language impairment (Pavelko, Owens, Ireland, and Hahs-Vaughn, under review). The use of vocabulary tests for diagnostic purposes is not well supported (Gray, Plante, Vance, & Henrichsen, 1999). This evidence suggests that clinicians may not be using data when selecting standardized tests.

Thus, existing research highlights the need for increased utilization of evidence-based practices in assessment of children with possible language disorders. The current study serves to examine whether frequently used child language assessments have psychometric adequacy and diagnostic accuracy, as well as to educate SLPs on evidence-based practices when selecting standardized assessments.

## Methods

### Selecting Assessment Tools for Review

A survey of 1,399 school-based SLPs reported the most frequently administered language assessment tools for the purpose of diagnosing children with language impairment (Pavelko, Ireland, Owens, & Hahs-Vaughn, under review). This nation-wide survey identified the following language assessment tools as most frequently used:

- *Clinical evaluation of language fundamentals*, 5th edition (CELF-5; Semel, Wiig, and Secord, 2013);
- *Comprehensive assessment of spoken language* (CASL; Carrow-Woolfolk, 1999);
- *Expressive one word picture vocabulary test*, 4<sup>th</sup> edition (EOWPVT-4; Martin and Brownell, 2010);
- *Oral and written language scales*, 2<sup>nd</sup> edition (OWLS-II; Carrow-Woolfolk, 2011);
- *Peabody picture vocabulary test*, 4<sup>th</sup> edition (PPVT-4; Dunn and Dunn, 2007);
- *Preschool language scale*, 5<sup>th</sup> edition (PLS-5; Zimmerman, Steiner, and Pond, 2011);
- *Receptive one word picture vocabulary test*, 4<sup>th</sup> edition (ROWPVT-4; Martin and Brownell, 2010);
- (ROWPVT-4; Brownell, 2010);
- *Test of language development: Intermediate*, 4<sup>th</sup> edition (TOLD-I:4; Newcomer and Hammill, 2008);
- *Test of language development: Primary*, 4<sup>th</sup> edition (TOLD-P:4; Newcomer and Hammill, 2008).

These nine assessment tools were selected to undergo review for psychometric adequacy and diagnostic accuracy. To initiate the review, the author of this study checked out each

assessment tool from the library at James Madison University or obtained the measure through personal connection. In addition, the *Assessment of literacy and language* (ALL; Lombardino, Lieberman, and Brown, 2005) and the *Test for examining expressive morphology* (TEEM; Shipley, Stone, and Sue, 1983) were selected for review because their diagnostic accuracy has been reported in the literature.

### **Procedures for Review of Selected Assessments**

Using McCauley and Swisher's (1984) recommendations, the researchers of the present study reviewed the examiner's manuals of each language assessment to determine whether it reported adequate psychometric properties. The following 10 psychometric criteria were applied to evaluate the psychometric characteristics of each assessment tool:

1. A clearly defined standardized sample, including: (a) Geographic residence, (b) Socioeconomic status, and (c) "Normalcy" of subjects.
2. A sample size of 100 or more.
3. Evidence of item analysis.
4. Measures of central tendency (mean and standard deviation).
5. Concurrent validity.
6. Predictive validity.
7. Test-retest reliability above .90.
8. Inter-examiner reliability above .90.
9. Test administration procedures.
10. Examiner qualifications.

The researchers of this study recorded all data in a Microsoft Excel spreadsheet. If the examiner's manual referenced the psychometric criteria the researchers noted it as present. In the

case that an examiner's manual failed to mention the criteria, the researchers noted it as Not Reported (NR). In addition to McCauley and Swisher's (1984) recommendations on psychometric criteria, the researchers reviewed each examiner's manual for purpose of the assessment, dialect considerations, language areas, and literacy areas. Lastly, the researchers analyzed the examiner's manual of each assessment to determine whether it reported data on diagnostic accuracy and whether sensitivity and specificity levels appeared acceptable, as described by Plante and Vance (1994).

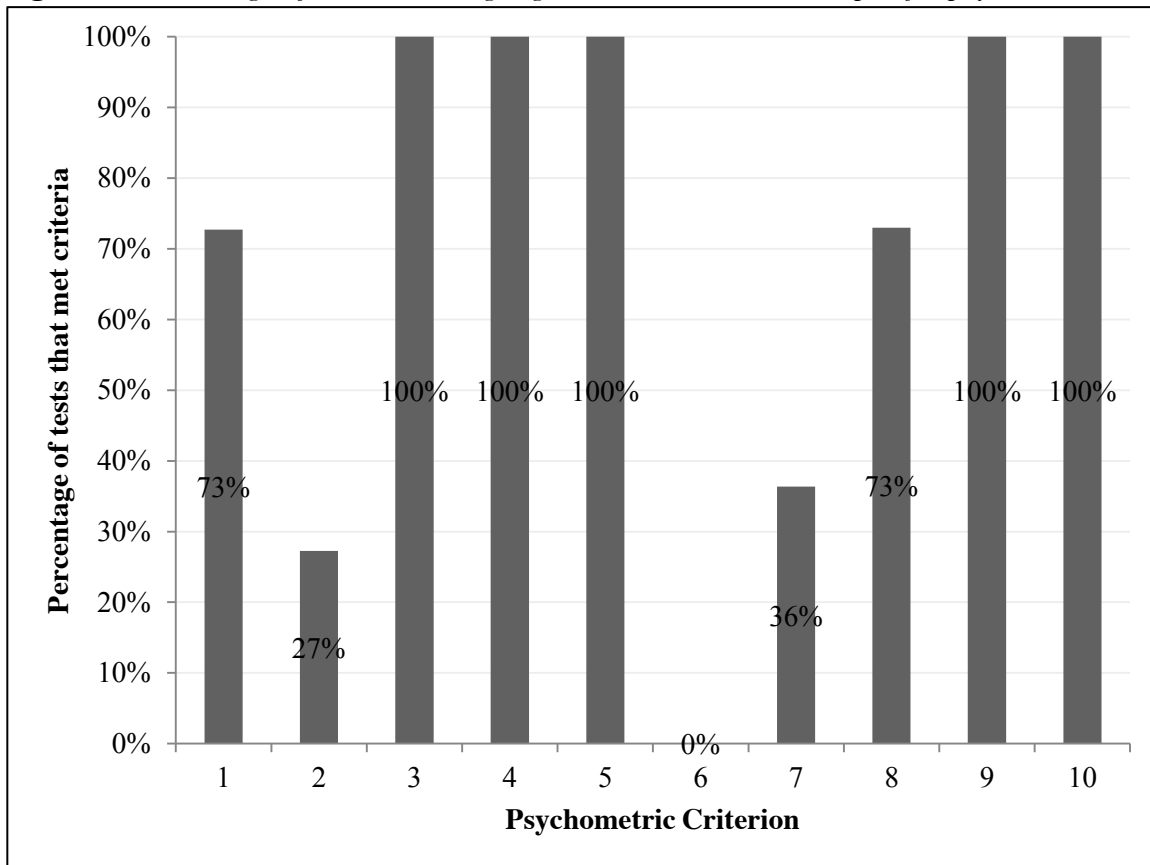
In the event that an examiner's manual did not provide information on diagnostic accuracy, the researchers searched the following databases to ascertain whether the data was available in published literature: PsycINFO – American Psychological Association, PubMed, Education Resources Information Center (ERIC), Cumulative Index to Nursing and Allied Health (CINAHL), and Mental Measurement Yearbook. Databases were searched in two ways. First, the researchers conducted a *Title* search by using the full name of the language assessment. If the researchers did not retrieve any results, the researchers conducted a *Title* search using the full name of the language assessment and specified the search to *Abstract AND Title*. The researchers reviewed the abstracts to determine whether articles provided information on the sensitivity and specificity of those assessment measures.

## **Findings**

### **Psychometric Characteristics**

Of the 11 reviewed language assessments, no one test met all 10 psychometric criteria, and results ranged from six to eight criteria met. Four tests met eight out of 10 psychometric criteria, five tests met seven out of 10 psychometric criteria, and two tests met six out of 10 psychometric criteria. Figure 1 presents a graphical representation of these findings. All assessments met the following criteria: item analysis, measures of central tendency (mean and standard deviation), concurrent validity, test administration procedures, and examiner qualifications. Nine tests clearly defined the standardized sample by providing information on (a) geographic residence, (b) socioeconomic status, and (c) the “normalcy” of subjects in the sample. The two tests that did not satisfactorily meet this criterion provided only two of the three required pieces of information about the standardized sample. None of the reviewed tests met the predictive validity criterion. Table 1 presents each test and whether the measure met each of the criteria.

**Figure 1.** Percentage of reviewed language assessments that met specific psychometric criterion.



*Note.* Psychometric criteria acquired from McCauley and Swisher (1984).

**Legend**

Number	Psychometric criterion
1	Clear standardized sample: (a) geographic residence (b) socioeconomic status (c) “normalcy” of subjects
2	Sample size $\geq 100$
3	Item analysis
4	Measures of central tendency
5	Concurrent validity
6	Predictive validity
7	Test-retest reliability $>.90$
8	Interexaminer reliability $>.90$
9	Administration procedures
10	Examiner qualifications

**Table 1.** *Presence or absence of psychometric criteria in reviewed language assessments.*

<b>Criteria</b>	<b>ALL</b>	<b>PPVT-4</b>	<b>TOLD-P:4</b>	<b>OWLS-II</b>	<b>TOLD-I:4</b>	<b>PLS-5</b>	<b>CASL</b>	<b>CELF-5</b>	<b>TEEM</b>	<b>EOWPVT-4</b>	<b>ROWPVT-4</b>
1. Clearly defined standardized sample:	•	•	•	•	•	•	•	•	•	•	•
(a) Geographic residence	•	•	•	•	•	•	•	•	•	•	•
(b) Socioeconomic status	•	•	•	—	•	•	•	•	•	•	•
(c) “Normalcy” of subjects	•	•	•	—	•	•	•	•	•	•	•
2. Sample size $\geq 100$	•	—	—	—	—	—	•	•	—	—	—
3. Item analysis	•	•	•	•	•	•	•	•	•	•	•
4. Measures of central tendency	•	•	•	•	•	•	•	•	•	•	•
5. Concurrent validity	•	•	•	•	•	•	•	•	•	•	•
6. Predictive validity	—	—	—	—	—	—	—	—	—	—	—
7. Test-retest reliability $> .90$	—	•	—	—	—	—	—	—	•	•	•
8. Inter-examiner reliability $> .90$	•	—	—	•	•	•	—	•	•	•	•
9. Test administration procedures	•	•	•	•	•	•	•	•	•	•	•
10. Examiner qualifications	•	•	•	•	•	•	•	•	•	•	•
<b>Total criteria met:</b>	<b>8/10</b>	<b>7/10</b>	<b>6/10</b>	<b>6/10</b>	<b>7/10</b>	<b>7/10</b>	<b>7/10</b>	<b>8/10</b>	<b>7/10</b>	<b>8/10</b>	<b>8/10</b>

## **Diagnostic Accuracy**

Of the 11 language assessments evaluated, six tests had published data regarding its diagnostic accuracy. Four of the six assessments with published diagnostic accuracy had sensitivity and specificity levels between the values of .80 and .90. These diagnostic levels would be considered “fair” in distinguishing the difference between children with language impairment and children with normal language development, based on Plante and Vance’s (1994) guidelines. The remaining five assessments provided no information on levels of sensitivity and specificity in its examiner’s manual or in the literature.

For two language assessment tools, the results found in the examiner’s manual differed from those of a third party review. According to a third party review, called the LEADERS project, both the PLS-5 and the CELF-5 failed to provide acceptable levels of diagnostic accuracy (LEADERS Project, 2013; LEADERS Project 2014). Table 2 summarizes the data on the diagnostic accuracy that each language assessment tool provided and indicates the conflicting results.

In addition, refer to Table 3 to compare and contrast psychometric adequacy and diagnostic accuracy across the eleven reviewed standardized assessments of child language. This representation indicates which tests that met a relatively large amount of psychometric criteria, tests that published diagnostic accuracy, and tests with acceptable levels of diagnostic accuracy.



**Table 2.** Tests with published and unpublished data on sensitivity and specificity.

<b>Test</b>	<b>Published data available</b>	<b>Sensitivity and Specificity .80-.90</b>
<b>ALL</b>	√	√
<b>PPVT-4</b>	X	X
<b>TOLD-P:4</b>	√	X
<b>OWLS-II</b>	X	X
<b>TOLD-I:4</b>	√	X
<b>PLS-5</b>	√	√ (Examiner's manual) X (LEADERS, 2013) <sup>a</sup>
<b>CASL</b>	X	X
<b>CELF-5</b>	√	√ (Examiner's manual) X (LEADERS, 2014) <sup>b</sup>
<b>TEEM</b>	√	√
<b>EOWPVT-4</b>	X	X
<b>ROWPVT-4</b>	X	X

<sup>a</sup> LEADERS Project (November 25, 2013). *Test Review: PLS-5 English*. Retrieved from <http://leadersproject.org/sites/default/files/PLS5-English-finaldraft.pdf>.

<sup>b</sup> LEADERS Project (February 17, 2014). *Test Review: CELF-5*. Retrieved from <http://leadersproject.org/sites/default/files/CELF5%20Test%20Review-LEADERS.pdf>.

**Table 3. Psychometric adequacy and diagnostic accuracy across reviewed language assessments.**

	ALL	PPVT-4	TOLD-P:4	OWLS-II	TOLD-I:4	PLS-5	CASL	CELF-5	TEEM	EOWPVT-4	ROWPVT-4
Number of psychometric criteria met?	8/10	7/10	6/10	6/10	7/10	7/10	7/10	8/10	7/10	8/10	8/10
Published diagnostic accuracy?	√	X	√	X	√	√	X	√	√	X	X
Acceptable diagnostic accuracy?	√	X	X	X	X	√, X <sup>a</sup>	—	√, X <sup>b</sup>	√	X	X

*Note.* Acceptable diagnostic accuracy according to recommendations by Plante and Vance (1994).

<sup>a</sup> LEADERS Project (November 25, 2013). *Test Review: PLS-5 English*. Retrieved from <http://leadersproject.org/sites/default/files/PLS5-English-finaldraft.pdf>.

<sup>b</sup> LEADERS Project (February 17, 2014). *Test Review: CELF-5*. Retrieved from <http://leadersproject.org/sites/default/files/CELF-5%20Test%20Review-LEADERS.pdf>.

## **Discussion and Clinical Implications**

The clinical decision-making process in speech-language pathology is an essential component of the practice and is critical for providing high quality services. However, the challenge of selecting assessments to use for diagnosing impairments in children may not always present itself as an easy task. An array of standardized tests exists for clinicians to choose from, each with their own strengths and weaknesses. To become informed users of the tool, SLPs must carefully evaluate the tests for both psychometric properties and diagnostic accuracy (Dollaghan, 2004). In order for professional practice to evolve within the field of speech-language pathology, SLPs must use data-driven decisions when selecting standardized tests for use as a part of a diagnostic evaluation.

The results of the current study, when compared to the results of previous research, reveal that improvements in the quality of standardized assessments have occurred with time. To date, more assessments offer empirical evidence that they are valid and can be used reliably to provide information concerning the existence of language impairment. As a whole, all of the reviewed assessment tools met more psychometric criteria than reviews in the past; however, many still lacked supporting data on diagnostic accuracy. The overall trend that emerges from this review, although not completely satisfactory, is encouraging.

To illustrate improved standardized assessment quality, in McCauley and Swisher's (1984) review, less than 20% of tests met five of the 10 psychometric criteria that the authors had employed. Then, using similar methodology, Plante and Vance's (1994) review indicated that 38% of the reviewed tests met at least five criteria. This points to a modest overall improvement in test validity when compared to the results of McCauley and Swisher's review (1984). Plante and Vance further analyzed the validity of standardized assessments by looking at diagnostic

accuracy (Plante & Vance, 1994; 1999). From this, the relatively stringent guideline for accurate discrimination, which states that 90% accuracy should be considered good, persists as a standard for clinical use (Plante & Vance, 1994). Today, much more information on test validity and reliability exists than in the past. The findings of the current study indicate that the overall psychometric validity of standardized assessments has improved. The 11 language assessments reviewed in this study met at least 60% of psychometric criteria, a significant step up from the 38% in 1994 and the 20% in 1984.

A trend to note from this study is that all of the reviewed assessment tools tend to meet those psychometric criteria that do not require extensive data collection, time to incorporate in examiner's manuals, or large amounts of financial resources, such as money. This trend confirms a similar observation from that of research in the past (McCauley & Swisher, 1984). Those assessments that met a high number of psychometric criteria and also published acceptable levels of diagnostic accuracy may have allocated more time, and/or financial resources, to collecting data. Authors of those assessments may also be more concerned or informed about the importance of such empirical evidence for use in diagnostics.

Item analysis, measures of central tendency (mean and standard deviation), and concurrent validity were met by 100% of the tests, which shows a marked improvement from McCauley and Swisher's study in 1984. This may have occurred due to increased awareness and promotion of systematic item analysis during item construction and/or selection (McCauley & Swisher, 1984). In addition, measures of central tendency and variance in test scores give users multiple ways of presenting the norms because it promotes flexibility in the use and interpretation of scores (Friberg, 2010). When a test meets the criterion for concurrent validity, it provides evidence that a correlation exists between results obtained from a given assessment tool

and an already valid measure that assesses a similar construct; a test is more likely to be valid if the two measures yield parallel results (Friberg, 2010). One hundred percent of the tests also met the criteria for test administration procedures and test user qualifications. As McCauley and Swisher (1984) stated, these two criteria require the least amount of financial resources and expenditure of time. Interestingly, not one of the assessments met the criteria of predictive validity. McCauley and Swisher (1984) state that to pass this criterion, test manuals need to include empirical evidence that it could be used to predict later performance. Predictive validity requires longitudinal studies in order to gather data; therefore, it is likely that time restrictions explain the meager data published. Regardless, it can be inferred from the current study that, as a whole, more standardized assessments meet greater numbers of psychometric criteria.

In regards to diagnostic accuracy, the current study highlights the need for improvement, as only four of the 11 assessments published data that indicated they could acceptably and accurately discriminate between children with language impairment and children with typically developing language. Unfortunately, research reveals that clinicians are selecting and administering assessment tools that either have poor or unreported diagnostic accuracy (Betz, Eickhoff, & Sullivan, 2013).

As mentioned earlier, the sensitivity and specificity values of a given assessment must be high enough to accurately discriminate between normal and impaired language (Gray, Plante, Vance, & Henrichsen, 1999). Without this type of empirical evidence, it is unknown whether the assessment can be used reliably and validly for the purpose of diagnosing children as language impaired. The type of assessment that a clinician selects impacts all children who undergo assessment, whether language impaired or not. A measure with poor diagnostic accuracy means that a certain number of children with language impairment will be incorrectly identified as

typically developing. Therefore, these children are unlikely to receive the interventions they need. Poor diagnostic accuracy also means that a certain number of children with normal language will be incorrectly identified as impaired. Therefore, these children could inappropriately be labeled as “disabled” and receive unnecessary services. Impairment categorizations influence service provision in both *who* gets served and the *quality* of those services. This should emphasize the need for each clinician to consider the social and educational consequences of under- or over- identification of language impairment on both the child who is impaired and the normally developing child who is misidentified (Gray, Plante, Vance, & Henrichsen, 1999).

The responsibility to become an educated and informed user of assessment tools used in diagnostics falls on clinicians themselves. McCauley and Swisher (1984) suggest that test users make themselves more aware of the psychometric properties and potential flaws within a test they consider using. This would help to reduce the impact of those flaws when coming to clinical decisions (McCauley & Swisher, 1984). If assessments display flaws, McCauley and Swisher (1984) recommend placing less weight on the results of the tests in the decision process and more on other objective and subjective evidence. Many states encourage SLPs to include a variety of measures, both standardized and non-standardized when conducting a comprehensive language assessment (Virginia Department of Education, 2011). Test users can also take advantage of their influence as a consumer and conduct careful psychometric reviews prior to test purchase. This practice would lead test authors and publishers to gather this empirical evidence in order to keep up with competing tests (and to reap the benefits of the purportedly difficult and expensive steps of test production).

This study does not intend to classify certain assessment tools as better or worse and encourage the use of certain assessments over others. Rather, this study serves solely to inform SLPs on evidence-based practices when it comes to selecting standardized assessments and to open discussion on its clinical implications. These findings are critical for any practicing SLP that administers standardized language assessments as a component of their diagnostic evaluations. Clinicians should use these findings, along with other recent research on this issue, as a form of guidance when selecting which assessments to administer during diagnostics.

## References

- American Speech-Language-Hearing Association. (2010). *Roles and responsibilities of speech-language pathologists in schools*. Available from [www.asha.org/SLP/schools/prof-consult/guidelines/](http://www.asha.org/SLP/schools/prof-consult/guidelines/).
- Betz, S., Eickhoff, J., & Sullivan, S. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*, 133-146.
- Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Torrance, CA: Western Psychological Services.
- Carrow-Woolfolk, E. (2011). *Oral and Written Language Scales (second edition)*. Torrance, CA: Western Psychological Services.
- Dollaghan, C. (2004). Evidence-based practice in communication disorders: what do we know, and when do we know it?. *Journal of Communication Disorders, 37*(5), 391-400.
- Dunn, L. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test (fourth edition)*. San Antonio, TX: Pearson Education.
- Friberg, J.C. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy, 26*, 77-92.
- Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30*(2), 196-206.
- Hammill, D. & Newcomer P. (2008). *Test of Language Development: Intermediate (fourth edition)*. Austin, TX: PRO-ED.



- Hammill, D. & Newcomer P. (2008). *Test of Language Development: Primary* (fourth edition). Austin, TX: PRO-ED.
- Hoffman, L.M., Loeb D.F., Brandel, J., Gilliam, R.B. (2011). Concurrent and construct validity of oral language measures with school-age children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 54, 1597-1608.
- LEADERS Project (November 25, 2013). *Test Review: PLS-5 English*. Retrieved from <http://leadersproject.org/sites/default/files/PLS5-English-finaldraft.pdf>.
- LEADERS Project (February 17, 2014). *Test Review: CELF-5*. Retrieved from <http://leadersproject.org/sites/default/files/CELF-5%20Test%20Review-LEADERS.pdf>.
- Lombardino, L., Lieberman, R., & Brown, J. (2005). *Assessment of Literacy and Language*. San Antonio, TX: Pearson Education.
- Martin A. & Brownell, R. (2010). *Expressive One Word Picture Vocabulary Test* (fourth edition). Torrance, CA: Western Psychological Services.
- Martin A. & Brownell, R. (2010). *Receptive One Word Picture Vocabulary Test* (fourth edition). Torrance, CA: Western Psychological Services.
- McCauley, R.J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, 49, 34-42.
- Merrell, A. & Plante, E. (1997). Norm-referenced test interpretation in the diagnostic process. *Language, Speech, and Hearing Services in Schools*, 28, 50-58.
- Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25, 15-24.

- Plante, E., & Vance, R. (1995). Diagnostic accuracy of two tests of preschool language. *American Journal of Speech-Language Pathology*, 4, 70–76.
- Semel E., Wiig E., & Secord, W. (2013). Clinical Evaluation of Language Fundamentals (fifth edition). San Antonio, TX: Pearson Education.
- ShIPLEY, K., Stone, T., & Sue, M. (1983). Test for Examining Expressive Morphology. Austin, TX: PRO-ED.
- Spaulding, T.J., Plante E., Farinella, K.A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37(1), 61-72.
- Virginia Department of Education, Office of Special Education and Student Services (2011). *Speech-Language Pathology Services In Schools: Guidelines For Best Practice*. Commonwealth of Virginia Department of Education. PDF file.
- Zimmerman, I., Steiner, V., & Pond, R. E. (2011). Preschool Language Scale (fifth edition). San Antonio, TX: Pearson Education.