

Florida State University Libraries

2019-12-01

The Student Oral Language Observation Matrix: A Psychometric Study With Preschoolers

Lindsay R. Dennis, S. Kathleen Krach, Michael P. McCreery and Sarah Navarro

The publisher's version of record is available at <https://doi.org/10.1177/1534508418782624>



Abstract

The Student Oral Language Observation Matrix (SOLOM) is an assessment of oral language skills. The aim of this study was to examine psychometric properties for the SOLOM for preschoolers through: (1) use of internal consistency methods to assess the reliability of the scores from the Student Oral Language Observation Matrix (SOLOM), and (2) examination of criterion-related validity by comparing the SOLOM with the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4; Dunn & Dunn, 2007). SOLOM internal consistency reliability was calculated using coefficient alpha; criterion-related validity was calculated using Pearson product moment correlation coefficients between the SOLOM and PPVT-4 scores of 90 preschool children. Results indicate that the SOLOM exhibits strong internal consistency reliability with criterion-related validity in the low to moderately correlated range. While additional research is necessary, results provide psychometric support for the use of the SOLOM as an informal assessment tool to be used by preschool teachers.

The Student Oral Language Observation Matrix (SOLOM):

A Psychometric Study with Preschoolers

The preschool years are a critical time for acquisition of both oral language as well as reading and reading-related skills. Results from a meta-analysis conducted by the National Early Literacy Panel (NELP) suggest that children's skills related to print knowledge (e.g., alphabet knowledge, print concepts), phonological processing skills (i.e., phonological awareness), and aspects of oral language (e.g., vocabulary, syntax/grammar, word knowledge) were substantive and independent predictors of children's later reading outcomes (Lonigan, Schatschneider, & Westberg, 2008a). Additionally, longitudinal studies indicate that children who are poor readers by the end of first grade almost never acquire average-level reading skills by the end of elementary school without sustained and substantial remediation efforts (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel, 1988; Torgeson & Burgess, 1998). Therefore, it is imperative that deficits in pre-reading skills, such as oral language, be identified as early as possible.

Oral language is the ability to use words to communicate ideas and thoughts and to use language as a tool to communicate with others (Dunst, Trivette, Masiello, Roper, & Robyak, 2006). Within the broader category of oral language, expressive language is the use of words to express meaning, and receptive language is the ability to listen, process, and understand the meaning of spoken words (Morgan & Meier, 2008). Of the skills developed in preschool, oral language is one of the most critical to be considered (Dickinson, Golinkoff, & Hirsh-Pasek, 2010), given its effect on reading achievement during both the early stages of learning to decode words, and the later stages of reading when the focus is on comprehension (NICHD Early Childhood Care Research Network, 2005; Storch & Whitehurst, 2002). A child's existing

vocabulary can predict his or her ability to learn new vocabulary, comprehend language, and access verbal information (Marchman & Fernald, 2008), all strong predictors of later reading. Therefore, oral language development appears to play a key role both in early literacy and later reading development (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Fernald & Weisleder, 2011).

Given that preschool children's performance on oral language tasks serve as a reliable indicator of their conventional literacy skill development, there is a need for psychometrically sound methods to identify children at-risk for later reading difficulties (Bradfield et al., 2014). A number of standardized, norm-referenced assessments exist for assessing young children's oral language development including the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4; Dunn & Dunn, 2007). Standardized assessments such as the PPVT-4 allow for meaningful comparisons among children because they have consistent administration and scoring criteria, provide scores that are reliable ($\alpha = .86$ to $.90$) and valid ($r = .79$), and generate standard scores from a normative sample. The importance of standardized assessments cannot be understated in terms of proper evaluation and identification of concerns with young children's early literacy development; however, an increasing selection of other oral language assessment measures are now available to assess emergent literacy skills (Lonigan, 2006a).

In their review of the evidence, Lonigan, Allan, and Lerner (2011) found that different methods and forms of assessments have specific strengths and weaknesses in terms of the types of the information yielded, and how that information can be used by teachers to provide instructional supports. Specifically, preschool teachers might conduct universal, periodic, emergent-literacy screening to find out what their children know and what they need to learn (Invernizzi, Landrum, Teichman, & Townsend, 2010). Additionally, quick, easy to use screening

measures involving simple classroom observation can be used to identify children in need of more formal diagnostic assessments (Lonigan et al., 2011). These types of observational assessments can include: checklists, ratings scales, or portfolios. For the purposes of this study, we focus on one informal teacher observational assessment, a rating scale, The Student Oral Language Observation Matrix (SOLOM).

The SOLOM (see Appendix 1) is intended to assess student's oral language skills. Skills are evaluated in the following categories: comprehension, fluency, vocabulary, pronunciation, and grammar (Brittain, 1991). There is no normative data nor are there age restrictions on the SOLOM. The authors state that the, "SOLOM is a rating scale that teachers can use to assess their students' command of oral language on the basis of what they observe" (Parker, Dolson, & Gold, 1985). Because it is in the public domain (and therefore free of charge), the SOLOM is a popular choice for assessing oral language ability (Saunders & O'Brien, 2006).

Although it may not have originally been designed as a language proficiency tool, the SOLOM has been adapted for use with students who are English language learners to determine language proficiency levels because it can be used as an evaluation of general language skills (Koran, 2015; Tucker, Donato & Antonek, 2008). Additionally, the SOLOM has been recommended for use as a language proficiency instrument by highly respected individuals in the field (Rhodes, Ochoa, & Ortiz, 2005; Saunders & O'Brien, 2006).

Despite the widespread use of the SOLOM, a comprehensive search found no psychometric studies published for this instrument. Therefore, additional data is needed for both reliability and validity for the SOLOM. It is beyond a single study to evaluate all aspects of reliability and validity within an instrument. Instead, the authors focused on two specific psychometric aspects of the SOLOM: internal consistency reliability and criterion-related

validity. The current study focused on preschoolers because oral language development is critical for this age group (Dickinson et al., 2010), and professionals need early screeners in order to intervene at the earliest possible time (Cabell, Justice, Zucker, Kilday, 2009).

Internal consistency reliability divides the instrument in parts (e.g., odd items and even items; first half vs second half of the instrument, etc.) and compares the scores on each part to the other (Sattler, 2008). Internal consistency data is a frequent method of calculating score reliability for instruments with multiple items. In this study, researchers used internal consistency methods to assess the reliability of the scores from the SOLOM.

The estimate of validity used by the researchers to assess the SOLOM is criterion-related validity. Criterion-related validity is a method of establishing psychometric similarity between one instrument, that measures a specific construct with a different instrument, purported to measure the same construct (Sattler, 2008). Often, this comparison is done using correlations between scores of an established measure and an untried one.

When asking how strong of a correlation is sufficient to determine psychometric validity, one of the best answers can be found in Messick's (1995) unified theory. This theory states that any validity findings must be considered within the context of: 1) what the test measures, 2) the purpose of the test, and 3) the consequences arising from the decisions made from test data. This is the criteria that the authors will use to determine the validity of the SOLOM.

The current article examines the criterion-related validity of the SOLOM by comparing it with the PPVT-4 (Dunn & Dunn, 2007). The PPVT-4 was chosen because it is a venerable measure of oral language that has a history of strong psychometrics for both validity and reliability (Dunn & Dunn, 2007; Kush & Shaw, 2010). It should be noted that the PPVT-4 measures only vocabulary and comprehension; whereas, the SOLOM measures other aspects of

oral language. Therefore, the researchers compared the PPVT-4 scores to each of the sections of the SOLOM as well as to the composite.

Methodology

Sample

Ten preschools were included in the study: four Head Start programs (n = 38), one religiously affiliated program (n = 14), and five preschools without a religious affiliation (n = 38). Only children whose parents consented to the IRB-approved data collection process were included. Out of the 90 total subjects, 48 were Caucasian/white, 36 were African American, one was Asian, three were multiracial, and two were “other.” Ethnicity information found that four preschoolers were Hispanic and 86 were not Hispanic. The children’s average age was 55 months (SD = 7.85 months), ranging from 39 months to 71 months of age. Out of the sample, 49 were male and 41 were female. Eighty-nine of the 90 households reported English was the language spoken at home, while one reported Spanish as the primary language.

Instruments

SOLOM. Teachers were asked to rate each child using the SOLOM (Parker et al., 1985). Children were rated on a scale of 1 (no proficiency) to 5 (fluent) across five domains (Comprehension, Fluency, Vocabulary, Pronunciation, & Grammar). These ratings were combined to form a composite language score. This composite score consisted of data related to both receptive and expressive language skills. The SOLOM is considered to be an indirect assessment because the teacher (not the child) provides the data.

PPVT-4. The children were administered the PPVT-4 (Dunn & Dunn, 2007) paper version, Form A. For the PPVT-4, the examiner directly assessed the child’s language skills by presenting stimuli and having the child respond. Specifically, the child was presented with an

array of four pictures. The examiner stated a single word; the child pointed to a picture represented by this word. Therefore, the PPVT-4 is a measure of receptive oral vocabulary and comprehension (Dunn & Dunn, 2007). The PPVT-4 measures two of the five aspects of oral language skills (i.e., comprehension and vocabulary) measured by the SOLOM.

The technical manual provides evidence supporting the reliability of the PPVT-4 with coefficients ranging from .86 to .90 for the ages used in the current study (Dunn & Dunn, 2007). Convergent validity data between the PPVT-4 and another test of language, the *Clinical Evaluation of Language Fundamentals, Fourth Edition* (CELF-4; Semel, Wiig, & Secord, 2003), produced “moderate-to-high” (Dunn & Dunn, 2007, p. 61) validity coefficients ($r = .77$ and $r^{adj} = .79$).

Procedures

The PPVT-4 was administered by school psychology graduate students who had completed coursework in testing and measurement. In addition, these graduate students were specifically trained by a licensed psychologist on standardization and administration of the PPVT-4. Scoring of the PPVT-4 protocols was done by a single, advanced school psychology student under supervision of a licensed psychologist. Prior to testing, teachers were asked to rate children’s language skills using the SOLOM. All ratings on the SOLOM and testing on the PPVT-4 were done within one month of each other between October and May (excluding December).

Data Analysis

Statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 23. Descriptive statistics for means and standard deviations (Table 1) were calculated for the PPVT-4: Form-A standard scores and the following scores on the SOLOM: Comprehension, Fluency,

Vocabulary, Pronunciation, and Grammar. A SOLOM Total score was also calculated by adding each area score; the descriptive statistics for this Total score were also calculated. Pearson product-moment correlation coefficients were calculated between each of the items of the SOLOM as well as the SOLOM Total Score against the PPVT-4: Form A. Reliability data was determined through the use of coefficient alpha.

Results

Descriptive statistics. Means and standard deviations for the SOLOM Total ($M = 20.83$; $SD = 4.63$), Comprehension ($M = 4.27$; $SD = .922$), Vocabulary ($M = 4.24$; $SD = .95$), and PPVT-4 ($M = 105.21$; $SD = 17.16$) are reported here (see Table 1). Table 2 provides the correlation matrix for the PPVT-4 and the SOLOM.

Validity. Pearson product-moment correlation coefficients between the SOLOM and PPVT-4 (Form A) indicated significant, moderate correlations between each of the SOLOM subscales and total score, and the PPVT-4 (see Table 2).

Reliability. In addition to criterion-related validity data, the authors calculated internal consistency reliability for the SOLOM. The coefficient alpha for the SOLOM fell in the high range ($\alpha = .946$).

Discussion

The preschool years represent a critical time for development of foundational early literacy skills associated with better reading outcomes (Lonigan, 2006a; Storch & Whitehurst, 2002). Early identification and intervention of young children struggling with early literacy skill development is critical; therefore, educators must be able to effectively assess an individual's strengths and needs (Cabell et al., 2009). Data from screening assessments can help educators

make decisions regarding the development of optimal teaching experiences (Invernizzi et al, 2010) or effective early literacy interventions (Cabell et al., 2009).

Standardized tests remain the most common approach to assess young children's emergent literacy skills for both screening and diagnostic purposes (Lonigan, 2006b). However, informal assessments can also be utilized to supplement or complement standardized tools as they offer several distinct benefits including: (a) efficiency in terms of time to administer, (b) relatively low in cost, and (c) may offer more in-depth developmental information (Cabell et al., 2009). Although well-respected in the field (Rhodes, et al., 2005; Saunders & O'Brien, 2006), the SOLOM has very little psychometric research to support its use as a potential informal assessment tool for measuring oral language skills. In particular, there was no research found on the instrument particular to preschool-aged children.

The results of this study suggest that the SOLOM has the potential to be a useful, informal screening instrument for preschool children's oral language skills. The results support previous research on the importance of early identification of children at-risk for later reading difficulties (Bradford et al., 2014) through assessment practices that provide teachers with the information needed to provide optimal learning experiences for children (Lonigan et al., 2011). Teacher report regarding children's emergent literacy skills, including oral language, have important implications in terms of distinguishing children who are more equipped with the skills necessary to become a proficient reader, from those who may require more support. This study adds an important contribution to the screening literature as the SOLOM can equip teachers to informally assess preschool children's skills, as well as provide them with information necessary to make valid educational decisions about designing clinical interventions (Cabell et al., 2009), or the need for further diagnostic assessment (Lonigan et al., 2011).

Further, the current study demonstrated that the SOLOM exhibits strong internal consistency reliability (Murphy & Davidshofer, 2004), when used with a preschool population. However, internal consistency is only one possible method of examining reliability. The researchers recommend that future studies focus on other forms of score reliability such as test-retest reliability, and/or inter-rater reliability to determine if the positive findings from this study continue.

In addition to score reliability, the current study also addressed criterion-related validity. The need for criterion-related validity was addressed comprehensively in the current *Standards for Educational and Psychological Testing*, which was published as a joint effort by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) in 2014. However, although the *Standards* describe how to conduct a criterion-related validity study, they did not establish the metric for interpreting any findings related to the results of these calculations. Specifically, they do not set a “cut-off” correlation coefficient expected between the two instruments. Professional judgement is needed to make that decision. Messick (1987) provided a method for using professional judgement when interpreting validity coefficients. He stated that any validity coefficient is made up of three sources of variance: actual differences in test performance, differences due to task expectations, and differences due to error. Ideally, the desired goal is to present a coefficient that represents mostly differences in test performance. Realistically, however, the coefficient will also consist of the other two sources of variability as well.

Differences due to error can be estimated by examining any known reliability data for the instrument. This is because reliability data (theoretically) consists of a ratio of actual differences

in test performance and test error (Sattler, 2008). Therefore, the reliability coefficient can be used as a metric for determining the amount of error, and thus the highest value of expected criterion-related validity. Because the SOLOM has a reliability coefficient of .95 from the current study and the PPVT-4 has a reliability coefficient as low as .86 from the manual (Dunn & Dunn, 2007), then theoretically the highest criterion-related validity coefficient possible would be .86. If it were any higher, the criterion-related validity results would be inconsistent with Messick's metrics (1987).

Differences due to tasks expectations must also be taken into consideration when considering a validity coefficient. Given that the task expectations between the SOLOM (indirect rating scale) and PPVT-4 differ (direct assessment), these differences may influence the coefficient more than if the comparison was made between two similar testing methods.

Previous research comparing direct and indirect assessments of functional ability have found statistically significant differences (Karagiozis, Gray, Sacco, Shapiro, & Kawas, 1998). Studies on direct and indirect assessments of children's physical ability resulted in correlation coefficients ranging from -0.56 to 0.89 , which were described as low to moderately correlated by the authors of that study (Adamo, Prince, Tricco, Connor-Gorber, & Tremblay, 2009). Although this is not a study of direct and indirect assessments for preschool language, these findings indicate that there may be great differences when comparing data from direct and indirect sources. These differences should be considered when evaluating the criterion-related validity coefficient between the SOLOM and the PPVT-4 of $r = .558$ for the SOLOM Total ($r = .468$ for Comprehension and $r = .528$ for Vocabulary), which were all found to be statistically significant at the 0.01 level. When considering both the error due to reliability and the error due

to data collection methodology, these coefficients should be considered as falling in the moderately correlated range.

To interpret the moderate validity coefficient, the researchers referred back to the questions found in Messick's (1995) unified theory. As a reminder, the theory states that any SOLOM validity findings must be considered within the context of: 1) what the test measures (oral language), 2) the purpose of the test (screening), and 3) the consequences arising from the decisions made from test data (low stakes decisions). Given these considerations, the moderate coefficient should be sufficient to indicate sufficient criterion-related validity for the total score on the SOLOM with the PPVT-4. However, when using Messick's (1995) standards, the moderate coefficient on the Vocabulary subscale on the SOLOM seems surprisingly low ($r = .528$). This would be considered low because the Vocabulary subtest and the PPVT-4 both measure the same construct "vocabulary." However, once you consider that the PPVT-4 measures receptive vocabulary and the SOLOM measures expressive vocabulary, the moderate coefficient may be less troubling. As a reminder, this validity data is only supportive when the SOLOM is used as a screening instrument. There is not sufficient evidence here to demonstrate validity of its use as a diagnostic instrument. Combined with the internal consistency reliability findings, the SOLOM appears to have psychometric evidence to support its use with a preschool population.

A few weaknesses in the current study should be addressed in this discussion. The largest weakness is that the researchers only assessed one aspect of reliability and validity as well as using only one age group. Additional studies should be conducted to examine other psychometric considerations as well as other age groups. One additional note is that although the SOLOM was originally designed as a screener for oral language skills, it has evolved to be an instrument

frequently used for evaluating language skills in English language learners. Since the current sample consisted almost exclusively of native English speakers, follow-up studies may need to be conducted for use of the SOLOM with a more linguistically diverse sample.

In sum, the results of this study provide information to support the use of the SOLOM as a potential informal assessment tool that be used by preschool teachers. Keeping in mind that screening tools like the SOLOM are meant to be used in combination with other standardized, norm-referenced assessments for the purposes of making educational decisions for young children. The SOLOM may provide important information to teachers and other stakeholders interested in early identification of potential concerns with children's early literacy development, when considering potential treatments and/or early intervention.

References

- Adamo, K. B., Prince, S. A., Tricco, A. C., Connor-Gorber, S., & Tremblay, M. (2009). A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: A systematic review. *International Journal of Pediatric Obesity*, 4(1), 2-27. doi: 10.1080/17477160802315010
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bradfield, T. A., Besner, A. C., Wackerle-Hollman, A. K., Albano, A. D., Rodriguez, M. C., & McConnell, S. R. (2014). Redefining individual growth and development indicators: Oral Language. *Assessment for Effective Intervention*, 39(4), 233-244. doi: 10.1177/1534508413496837
- Brittain, F. P. (1991). Effects of a maintenance bilingual bicultural program on fully English proficient students. *The Journal of Educational Issues of Language Minority Students*, 8, 125-146.
- Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, & Hearing Services in Schools*, 40(2), 161-173. doi: 10.1044/0161-1461(2009/07-0099)
- Dickinson, D., Golinkoff, R., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, 39, 305-310. doi:10.3102/0013189X10370204
- Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships

- among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology*, *95*, 465–481. doi:10.1037/0022-0663.95.3.465
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Bloomington, MN: NCS Pearson, Inc.
- Dunst, C., Trivette, C. M., Masiello, T., Roper, N., & Robyak, A. (2006). Framework for developing evidence-based early literacy learning practices. *CELL papers*, *1*, Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v1_n1.pdf
- Fernald, A., & Weisleder, A. (2011). Early language experience is vital to developing fluency in understanding. *Handbook of Early Literacy Research*, *3*, 3–19
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal individual growth curve analysis. *Journal of Educational Psychology*, *88*, 3 – 17. doi: 10.1037/0022-0663.88.1.3
- Invernizzi, M., Landrum, T.J., Teichman, A., & Townsend, M. (2010). Increased implementation of emergent literacy screening in pre-kindergarten. *Early Childhood Education Journal*, *37*, 437-446. doi: 10.1007/s10643-009-0371-7
- Karagiozis, H., Gray, S., Sacco, J., Shapiro, M., & Kawas, C. (1998). The Direct Assessment of Functional Abilities (DAFA): A comparison to an indirect measure of instrumental activities of daily living. *The Gerontologist*, *38*(1), 113-121. doi: 10.1093/geront/38.1.113
- Koran, E. (2015). Assessment of EFL learners' sociolinguistic and pragmatic competence and performance. *Journal of Education in Black Sea Region*, *1*, 45-53.

- Kush, J. C., & Shaw, S. (2010). Review Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4). In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook*. Lincoln, NE: The Buros Center for Testing.
- Lonigan, C. J. (2006a). Conceptualizing phonological processing skills in prereaders. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research: Volume 2* (pp. 77-89). New York: Guilford Press.
- Lonigan, C. J. (2006b). Development, assessment, and promotion of pre-literacy skills. *Early Education and Development, 17*(1), 91-114. doi: 10.1207/s15566935eed1701_5
- Lonigan, C. J., Allan, N. P., & Lerner, M. D. (2011). Assessment of preschool early literacy skills: Linking children's educational needs with empirically supported instructional activities. *Psychology in the Schools, 48*(5), 488-501. doi: 10.1002/pits.20569
- Lonigan, C. J., Schatschnieder, C., & Westberg, L. (2008a). Impact of code-focused interventions on young children's early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel* (pp. 107 – 151). Washington, DC: National Institute for Literacy.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, F9–F16. doi:10.1111/j.1467-7687.2008.00671.x
- Messick, S. (1987). Validity. *ETS Research Report Series, 2*, i-208. doi: 10.1002/j.2330-8516.1987.tb00244.x
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749. doi: 10.1037/0003-066X.50.9.741

- Morgan, P. L., & Meier, C. R. (2008). Dialogic reading's potential to improve children's emergent literacy skills and behavior. *Preventing School Failure, 52*, 11–16.
doi:10.3200/PSFL.52.4.11-16
- Murphy, K. R. & Davidshofer, C. O. (2004). *Psychological testing: Principles and applications (6th Edition)*. Upper Saddle River, NJ: Prentice Hall.
- NICHHD ECCRN. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*, 428–442. doi:10.1037/0012-1649.41.2.428.
- Parker, D., Dolson, D., & Gold, N. (1985). *Student Oral Language Survey (SOLOM)*. Sacramento, CA: Bilingual Education Office of the California Department of Education.
Retrieved from <http://www.cal.org/twi/EvalToolkit/appendix/solom.pdf>
- Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. NY: Guilford Press.
- Sattler, J. (2008). *Assessment of children: Cognitive foundations (5th ed.)*. San Diego, CA: Author.
- Saunders, W. M. & O'Brien, G. (2006). Oral language. In F. Genesee, K. Lindholm-Leary, W. Saunders, & D. Christian (Eds.), *Educating English language learners* (pp. 2-64). Cambridge, UK: Cambridge University Press.
- Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4)*. Bloomington, MN: NCS Pearson, Inc.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934-947.
doi: 10.1037/0012-1649.38.6.934

- Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161 – 188). Mahwah, NJ: Lawrence Erlbaum.
- Tucker, G. R., Donato, R., Antonek, J. L. (2008). Documenting growth in a Japanese FLES program. *Foreign Language Annals*, 29, 539-550. doi: 10.1111/j.1944-9720.1996.tb01265.x

Table 1

Means and Standard Deviations for SOLOM and PPVT-4

Instrument	N	M	SD
PPVT-4			
Form A: Paper-based	90	105.21	17.16
SOLOM			
Total	90	20.83	4.63
Comprehension	90	4.27	.922
Fluency	90	4.16	1.02
Vocabulary	90	4.24	.95
Pronunciation	90	4.06	1.17
Grammar	90	4.11	1.03

Table 2

Correlation Matrix for SOLOM and PPVT-4

	PPVT-4		SOLOM				
PPVT-4	Form A	Total	Comp.	Fluency	Vocab	Pronun	Gram
Form A: Paper-based		.558**	.468**	.530**	.528**	.492**	.521**
SOLOM							
Total	.558**		.871**	.937**	.935**	.885**	.925**
Comprehension (Comp.)	.468**	.871**		.831**	.796**	.635**	.748**
Fluency (Fluency)	.530**	.937**	.831**		.844**	.781**	.819**
Vocabulary (Vocab)	.528**	.935**	.796**	.844**		.778**	.852**
Pronunciation (Pronun)	.492**	.885**	.635**	.781**	.778**		.789**
Grammar (Gram)	.521**	.925**	.748**	.819**	.852**	.789**	.

** Correlation is significant at the 0.01 level (2-tailed).

Appendix 1: Student Oral Language Observation Matrix (SOLOM)

The SOLOM is not a test per se. A test is a set of structured tasks given in a standard way. The SOLOM is a rating scale that teachers can use to assess their students' command of oral language on the basis of what they observe on a continual basis in a variety of situations - class discussions, playground interactions, and encounters between classes. The teacher matches a student's language performance in a five mains - listening comprehension, vocabulary, fluency, grammar, and pronunciation - to descriptions on a five-point scale for each (See Figure 1). The scores for individual domains can be considered, or they can be combined into a total score with a range of five through 25, where approximately 19 or 20 can be considered proficient. SOLOM scores represent whether a student can participate in oral language tasks typically expected in the classroom at his or her grade level.

Because it describes a range of proficiency from non-proficient to fluent, the SOLOM can be used to track annual progress. This, in turn, can be used in program evaluation, and as some of the criteria for exit from alternative instructional programs. However, to be used for these purposes, it is important to ensure that all teachers who use it undergo reliability training so that scores are comparable across teachers. For this purpose, a training video has been produced by Montebello School District in California.

The SOLOM does not require a dedicated testing situation. To complete it, teachers simply need to know the criteria for the various ratings and observe their students' language practices with those criteria in mind. Therein lies the greatest value of the SOLOM and similar approaches:

- it fixes teachers' attention on language-development goals;
- it keeps them aware of how their students are progressing in relation to those goals; and
- it reminds them to set up oral-language-use situations that allow them to observe the student, as well as provide the students with language-development activities.

While observing, teachers should be attuned to the specific features of a student's speech that influenced their rating. They can use this information as a basis of instruction. The SOLOM is sufficiently generic to be applicable to other language besides English.

The SOLOM is not commercially published. It was originally developed by the San Jose Area Bilingual Consortium and has undergone revisions with leadership from the Bilingual Education Office of the California Department of Education. It is within the public domain and can be copied, modified, or adapted to meet local needs.

Directions for Administering the SOLOM:

Based on your observation of the student, indicate with an "X" across the category which best describes the student's abilities.

- The SOLOM should only be administered by persons who themselves score at level "4" or above in all categories in the language being assessed.
- Students scoring at level "1" in all categories can be said to have no proficiency in the language.

SOLOM Teacher Observation Student Oral Language Observation Matrix					
Student's Name:			Grade:	Date:	
Language Observed:			Administered By (signature):		
	1	2	3	4	5
A. Comprehension	Cannot be said to understand even simple conversation.	Has great difficulty following what is said. Can comprehend only social conversation spoken slowly and with frequent repetitions.	Understands most of what is said at slower-than-normal speed with repetitions.	Understands nearly everything at normal speech. Although occasional repetition may be necessary.	Understands everyday conversation and normal classroom discussions.
B. Fluency	Speech so halting and fragmentary as to make conversation virtually impossible.	Usually hesitant: often forced into silence by language limitations.	Speech in everyday conversation and classroom discussion frequently disrupted by the student's search for the correct manner of expression.	Speech in everyday conversation and classroom discussions generally fluent, with occasional lapses while the student searches for the correct manner of expression.	Speech in everyday conversation and classroom discussions fluent and effortless; approximating that of a native speaker.
C. Vocabulary	Vocabulary limitations so extreme as to make conversation virtually impossible.	Misuse of words and very limited: comprehension quite difficult.	Student frequently uses wrong words: conversation somewhat limited because of inadequate vocabulary.	Student occasionally uses inappropriate terms and/or must rephrase ideas because of lexical inadequacies.	Use of vocabulary and idioms approximate that of a native speaker.
D. Pronunciation	Pronunciation problems so severe as to make speech virtually unintelligible.	Very hard to understand because of pronunciation problems. Must frequently repeat in order to make him/herself understood.	Pronunciation problems necessitate concentration on the part of the listener and occasionally lead to misunderstanding.	Always intelligible, although the listener is conscious of a definite accent and occasional inappropriate intonation patterns.	Pronunciation and intonation approximate that of a native speaker.
E. Grammar	Errors in grammar and word order so severe as to make speech virtually unintelligible.	Grammar and word order errors make comprehension difficult. Must often rephrase and/or restrict him/herself to basic patterns.	Makes frequent errors of grammar and word order that occasionally obscure meaning.	Occasionally makes grammatical and/or word order errors that do not obscure meaning.	Grammar and word order approximate that of a native speaker.

Note: This is considered public domain and was obtained at <http://www.cal.org/twi/EvalToolkit/appendix/solom.pdf>