Early Learning Scale
Technical Report

Shannon Riley-Ayers, Ph.D.
Ellen C. Frede, Ph.D.
Kwanghee Jung, Ph.D.

National Institute for Early Education Research

September 2010

The Early Learning Scale (ELS; Riley-Ayers, Boyd, & Frede) is a systematic assessment for preschool children conceptually derived from the New Jersey Early Learning Assessment System- Language Arts Literacy (ELAS-L; Wolock et al., 2003) and Math (ELAS-M; Whelan, Boyd, & Frede, 2007). It is designed for teachers to assess children's progress toward learning standards such as the *Head Start Child Outcomes Framework* and the State Learning Expectations. This assessment is a concise, performance-based measure using student data collected through observation and work samples. Data is analyzed using research-based benchmarks and assigned a score on the 5-point continuum. This is a manageable system with 10 items that provide the teacher valuable data to inform instruction and improve student learning across domains. In addition, since this system is based on state early learning standards and current research and is not curriculum-specific, the ELS can be used in any classroom.

The National Institute for Early Education Research (NIEER) developed this observation-based performance assessment in response to a request by educators for a comprehensive, standards-based assessment system capable of informing instruction and making an impact on teaching and learning. Standardized tests are not an appropriate response to this request and often are misused. Teachers need a formative, on-going assessment tool for their classroom.

Performance-based assessments are necessary because children change from situation to situation and day to day. Performance assessments are able to capture children's skills and knowledge in real life over time. This type of assessment system also compares children to themselves, is comprehensive, and focuses on strengths and interests, which differs greatly from standardized tests. Lastly, and perhaps most important, this assessment approach informs teaching and can be easily used to communicate with parents in a meaningful way about their child's growth and development during the preschool years.

The Instrument

The ELS examines three domains with a total of 10 items across the domains. Within each item are strands that further delineate the items for more focused observation. The 5-point continuum has indicators at levels 1, 3, and 5. Scores are reported for each of the 10 items.

Domains and items that are included are measurable, develop on a continuum, and are critical to present and future learning. The ELS includes items in math/science, social emotional/social studies, and language arts literacy. The Early Learning Scale does not provide a continuum for the arts and physical development. At the preschool level, standards dictate children should begin to explore and develop an appreciation for the arts. However, a child's appreciation of something is difficult to observe and not appropriate to place on a continuum. Additionally, physical development is usually best assessed using a checklist rather than a continuum. The two domains are included because of their importance for teachers to notice and intervene when concerns arise. So, we provide the research base and a location to collect evidence, but these areas are not scored on the continuum.

Training and Support Approach

Comprehensive training and sustainability are the foundation for implementation of the ELS. Entities using the system are required to commit to intensive training. The training consists of an introduction to the system and assessment approach including focused observation and quality documentation. Child development, effective teaching strategies and current research permeates each training module and provides teachers the foundation for effectively utilizing this assessment to inform and improve instruction. Educators learn the system over time and practice using the instrument during training sessions and in the classroom, first in small pieces and then as a whole. Analyzing data, assigning scores, and planning instruction based on this information is presented and demonstrated. Training groups are kept small to ensure adequate opportunities for questions and discussion, small group work, and individualized support.

Individualized support is enhanced by the inclusion of coaches (master teachers) in the initial training. Another key component of the training is coach-led workgroups where a group of teachers comes together with their coach as facilitator to present and discuss student work. This provides the teachers the opportunity to seek support from their colleagues and engage in educational discussions regarding using assessment to drive instruction. Lastly, we encourage coaches to work with teachers in the classroom on an individual basis.

Teachers implementing the ELS are trained by the coaches or master teachers. NIEER representatives offer support and guidance. Work groups and classroom coaching will continue to be a necessary component of the implementation of the instrument.

Training modules are supported by the Guide Book, which accompanies the instrument and offers teachers detailed information about the system and about each domain and item. For each item, the Guide Book provides a research base, continuum descriptions, ideas for teaching and documenting, sample anecdotes, and a list of resources for further reading on the topic. The Guide Book also includes all the necessary forms for implementation including the anecdotal record forms, class record form, and child accomplishments summary form, which is used for communication with parents about the child's development.

A final component of the training approach is assessment of teachers' reliability on the instrument. Once the teachers have been trained with the instrument and implemented it in their classrooms for at least one score period (1/3 of the year), they are assessed on their scoring of data using the instrument. The process is outlined in more detail below. Teachers who reach 70 percent reliability are considered independent with the instrument, those between 60 and 69 percent need more intensive support beyond the work groups and will require more one-to-one coaching in their classrooms. Teachers scoring below 60 percent agreement need to be retrained.

Reliability

Inter-rater reliability was assessed for the ELS to determine teachers' reliability of scoring data using the ELS instrument.

Sample

Fifty-seven teachers in one large urban district in New Jersey participated in the training and reliability assessment for the ELS. They all hold a bachelor's degree or higher and have an average of four years teaching experience. Approximately half of the teachers had already been trained and used a performance-based assessment system before being trained with the ELS. Four trainers of ELS were also administered the reliability assessment.

A second group of educators were included in the reliability assessment. In South Dakota, three types of educators were trained with the ELS. Twenty-nine trainers learned the ELS as a part of South Dakota's Early Learning Guidelines (ELG) Trainers Training. Another cohort included seven prekindergarten teachers and two prekindergarten administrators who received training on the ELS along with the South Dakota's ELG. Lastly, 13 practicum students were trained in ELS as part of their practicum seminar.

A second cohort of teachers from a New Jersey school district were trained by NIEER and worked with the instrument. Reliability data then was collected. A total of 17 teachers completed the training and assessment.

Procedures

The first step for the reliability assessment was to generate six complete folios for the ELS assessment. These were collected from data in the field and collated to create six complete folios with sufficient data to score each of the 10 items. Then, experts in the field of early education and performance-based assessment reviewed and scored the folios. An agreed-upon score of 1-5 was determined through discussion and clarification of the evidence for each item on the six folios. This score is considered the true score for the item and teachers' exact agreement with the scores determine their reliability score.

In New Jersey, 57 teachers independently scored three folios (total 30 items across three children) directly following their initial training in ELS. The second collection of three folios (total 30 items across three children) was scored after 3-6 months of using the instrument in their classrooms. In South Dakota, the reliability assessment for all participants occurred directly following the ELS training. The third group of teachers scored three folios after training and one score period of use with the instrument.

Reliability Results

The average reliability for all 57 teachers in New Jersey across the two administrations of the reliability assessment is 71 percent. After item analysis of the reliability folios we noted that for six items (one from the first administration and five from the second administration) more than 40 percent of the teachers agreed that the score was different than the determined

true score. This indicates that there may be an issue with the item. When the average reliability is re-examined with these six items removed it increases to 77 percent. Similarly, for the four trainers who were assessed for reliability, the average reliability increases from 91 to 98 percent. See Table 1 below for further details of the breakdown of teacher reliability.

Table 1. NJ ELS Reliability Data

	All Items	Six Items Deleted
Average Reliability (N=57)	71%	77%
Teachers Below 60% *Need retraining*	9 (16%)	1 (2%)
Teachers between 60 and 68% *Need support*	16 (28%)	15 (26%)
Teachers at 70% or Higher *Independent*	32 (56%)	41 (72%)
Trainers' Reliability (N=4)	91%	98%

The average reliability for the three cohorts of educators in South Dakota ranged from 73.5 to 77.6 percent average agreement. South Dakota participants completed one reliability administration set of three folios for reliability. See Table 2 for further details.

Table 2. South Dakota ELS Reliability Data

Trainers (N=29)	Average Agreement 77.5% Range 53 – 93% 24 trainers at 70% or above
Pre-K Teachers and Administrator (N=9)	Average Agreement 76.8% Range 70 – 90% 9 teachers at 70% or above
Practicum Students (N=13)	Average Agreement 73.5% Range 53 – 97% 9 students at 70% or above

The average reliability for the final group of teachers was 80 percent on the administration of three folios. See Table 3 for further details.

Table 3. NJ Cohort 2 ELS Reliability

	Three folios (all items)
Average Reliability (N=17)	80%
Teachers Below 60% *Need retraining*	1 (6%)
Teachers between 60 and 68% *Need support*	1 (6%)
Teachers at 70% or Higher *Independent*	15 (88%)

Validity

Validity was first analyzed by looking at internal consistency. The Cronbach's alpha for the ELS instrument demonstrates high internal consistency at .91. Concurrent validity was also examined for the ELS by comparing it to established instruments. This research is described below.

Sample

Participants were 285 children from one district in NJ across 10 schools with 57 teachers. The classrooms were taught by the 57 teachers who were trained on the instrument, implemented it in their classrooms, and participated in the reliability component of the present study as NJ cohort 1.

Procedures

Five children were randomly chosen from each classroom to participate. The classroom teacher completed the ELS with the child by collecting data during the score period of November to February, evaluating the evidence and providing a score. Children were tested by NIEER assessors using the Early Literacy Skills Assessment (ELSA; DeBruin-Parecki, 2005) during one session and the Child Math Assessment (CMA; Klein & Starkey, 2006) on another day. These assessments took place between the beginning of February and mid-April.

These assessments best match the components of the ELS. A science instrument was not available for use and the social-emotional scales that were available at the time of the study

consisted of mainly teacher rating scales and would not be useful to establish concurrent validity for the ELS because that score is also determined by the teacher.

Measures

Early Literacy Skills Assessment (ELSA; DeBruin-Parecki, 2005)

The ELSA is a child assessment that measures four key principles of early literacy – comprehension, phonological awareness, alphabetic principle, and concepts about print. It has 23 items presented in a children's storybook form, which in our experience makes it very attractive to children. There are two protocols that are both available in Spanish and English. One of the major advantages of the ELSA is that it assesses a broad range of language and literacy constructs including comprehension, phonological awareness, alphabetic principle, and concepts about print. An extensive investigation of the ELSA's psychometric properties conducted by an outside evaluator found good reliability and validity. "Taken in sum, these results confirm the reliability of the ELSA as a measure of children's early literacy skills. Furthermore, the consistency of the results supports the general validity of the ELSA constructs for assessing both English and Spanish-speaking populations" (p. 9, Cheadle, 2007).

This instrument was specifically selected because of its close match to the ELS. The inclusion of comprehension made it particularly aligned. The exception is that writing is not assessed through the ELSA but is included as an item on the ELS.

Child Math Assessment (CMA; Klein & Starkey, 2004)

The CMA was selected because it measures nine key principles of early math – counting, one-set addition and subtraction, two-set addition and subtraction, geometric reasoning, construction of equivalent sets, direct measurement, shape recognition, pattern duplication, and division. It has nine tasks presented in a hands-on form using manipulatives, which in our experience makes it very attractive to children. There are two protocols that are both available in Spanish and English. One of the major advantages of the CMA is that it assesses a broad range of math constructs. An extensive investigation of the CMA's psychometric properties found good reliability and validity. "Test-retest reliability over a 14-day interval is .910, and Cronbach's alpha over all tasks is .898. In addition, we administered the TEMA-3 along with the CMA in order to obtain concurrent validity with another standardized measure of early number knowledge. We obtained significant correlations between the CMA Composite Score and the TEMA Math Ability Score (.741 - .748). This is consistent with our prediction that the CMA would correlate well, but not completely overlap, with the TEMA because the CMA assesses a broader range of informal mathematical knowledge than the TEMA" (A. Klein, personal communication, July 5, 2007).

Results

Concurrent validity was examined using partial correlations to control for date of test administration because the NIEER assessments using the ELSA and CMA spanned such a large time frame around the score date for the ELS. Correlations were generally significant and low to moderate.

Language Arts/Literacy

The total ELSA raw score and the total ELS Language Arts/Literacy score correlated moderately at .36. The correlations between the ELS items and the ELSA total range from .23 to .33. ELS item number 10, Writing, correlation with ELSA is the lowest because the ELSA does not examine writing. The ELSA subscales correlate with the total score for ELS Language Arts/Literacy domain from .16 to .37. Phonological awareness correlates the lowest, and we have seen in the field that this is the item on the ELS that teachers struggle most with understanding and collecting accurate and meaningful data for evaluation. The other low correlation is comprehension and we believe that comprehension often poses a difficulty in assessing. This perhaps is the reason that a comprehension assessment was difficult to find. The ELSA does not provide concurrent validity on the comprehension component of the instrument in its technical report specifically because no early literacy assessment measuring comprehension was located. See Table 4.

Table 4. Correlations of ELSA and ELS

	Total Raw ELSA	ELSA Comprehension	ELSA Phonological Awareness	ELSA Alphabetic Principal	ELSA Concepts of Print
Total ELS Lang. Arts	.36**	.19**	.16**	.37**	.31**
ELS 7: Oral Language	.28**	.16**	.14*	.33**	.22**
ELS 8: Phonological Awareness	.33**	.23**	.22**	.29**	.20**
ELS 9: Print Awareness	.33**	.14*	.14*	.34**	.31**
ELS 10: Writing	.23**	.08	.04	.24**	.26**

$*p<.05; **p<.01$

Math

The total Math/Science ELS domain and the CMA total score correlate significantly at .46. The correlations between the ELS Math/Science items and the CMA total range from .17 to .43. The lowest correlation is between ELS and the CMA subscales Equivalent Sets and Division. This is not surprising since these skills are not directly measured by the ELS. The highest correlation is between ELS Math/Science and CMA subscale counting, which is directly measured by the ELS. The correlations between the CMA total and the ELS Math/Science items range from .35 to .46, which are moderate. See Table 5.

Table 5. Correlations of CMA and ELS

	CMA Total	CMA Counting	CMA 1 Set Add/Sub	CMA Geometry	CMA Equiv. Sets	CMA 2 Set Add/Sub	CMA Measure-ment	CMA Shapes	CMA Patterns	CMA Division
Total ELS Math/ Science	.46**	.43**	.28**	.22**	.17*	.25**	.34**	.19**	.24**	.17**
1: Number	.42**	.47**	.28**	.15*	.13*	.21**	.36**	.16*	.22**	.12
2: Geometry	.38**	.34**	.16*	.19**	.18**	.22**	.24**	.17*	.19**	.18**
3: Observation	.35**	.32**	.21**	.20**	.15*	.16*	.21**	.18**	.22**	.10
4: Inquiry	.40**	.33**	.28**	.20**	.11	.27**	.31**	.15*	.19**	.19**

*p<.05; **p<.01

Conclusion

The psychometrics of the ELS are comparable to published instruments in the field of early childhood that use the same assessment approach.

We found average inter-rater reliability of teachers to range from .71 to .77, while our trainers were considerably higher at .91 to .98. High/Scope Child Observation Record (COR) reports inter-rater agreement at .69 to .73 and .93 for research assistants (High/Scope Educational Research Foundation, n.d.). The Work Sampling System reports .88 agreement between two outside coders, but .68 and .72 for teachers with the outside coders (Meisels, Liaw, Dorfman, & Nelson, 1995) and this research is on grades kindergarten through third, not with preschool. Lastly, the Developmental Continuum from Creative Curriculum does not report inter-rater reliability (Lambert, n.d.).

The correlations for the ELS to the standardized measures range from .39-.46 on whole instrument comparisons [the ELS to the Early Literacy Skills Assessment (ELSA; DeBruin-Parecki, 2005) and the Child Math Assessment (CMA; Klein & Starkey, 2004)] ; subscale relationships were lower again considering that the standardized tests, although closely aligned were not an exact match to what is assessed on the ELS (i.e., writing was not included in the standardized literacy assessment so there is no surprise that there was not a strong relationship between the writing subscale on the ELS and the ELSA total score). Other instruments report correlations between the instrument and standardized assessments. The COR correlated with the Cognitive Skills Assessment Battery in the range of .46-.62 (High/Scope Educational Research Foundation, n.d.) and the Work Sampling System reports .36-.75 correlated with the Woodcock-Johnson (Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001).

Note that more than half the teachers in the ELS validity study had only used the instrument for one score period prior and the rest of the participants were using the instrument for the first time. We are confident that with further use of the instrument the teachers' reliability will continue to increase, which will improve the concurrent validity of the instrument. Additionally, all of the content measured on the ELS cannot be directly correlated to standardized measures, thus lowering the subscale correlations. Further examination of the reliability and validity of the ELS will shed light onto these issues.

References

Cheadle, J. (2007). Early Literacy Skills Assessment psychometric report: For both English and Spanish versions. Ypsilanti, MI: High/Scope Press. Retrieved from http://www.highscope.org/file/Assessment/ELSAJacobs.pdf

DeBruin-Parecki, A. (2005). *Early Literacy Skills Assessment.* Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation (n.d.). Child Observation Record (COR): Information for decision makers. Retrieved from http://www.highscope.org/file/Assessment/cor_decision_maker.pdf.

Klein, A. & Starkey, P. (2006). *Child Math Assessment.* University of California, Berkeley.

Lambert, R. G. (n.d.). The Developmental Continuum Assessment System for Ages 3 to 5: The assessment component of The Creative Curriculum for Preschool. Retrieved from http://www.teachingstrategies.com/content/pageDocs/Dev_Continuum_Research_Report_1_04.pdf?CFID=4260110&CFTOKEN=80972503.

Meisels, S., Liaw, F., Dorfman, A., & Nelson, R. F. (1995). The Work Sampling System: Reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly, 10*, 277-296.

Meisels, S., Bickel, D., Nicholson, J., Xue, J., Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38(1)*, 74-95.

Riley-Ayers, S., Boyd, J. S., & Frede, E. (2008). Improving Teaching through Systematic Assessment: Early Learning Scale. NIEER: New Brunswick, NJ.

Whelan, R., Boyd, J. S., & Frede, E. (2007). *Early Learning Assessment System - Math.* Trenton, NJ: New Jersey Department of Education.

Wolock, E. (with The Early Learning Improvement Consortium). (2003). *Early Learning Assessment System - Language Arts Literacy*. Retrieved from http://www.nj.gov/education/ece/curriculum/elas/.