TECHNICAL REPORT OF THE COR FOR KINDERGARTEN VALIDATION STUDY

Technical Report of the

COR for Kindergarten
Validation Study

January 2016

Acknowledgments

We would like to extend our heartfelt appreciation to the supervisors, teachers, parents, and children who contributed their precious time to these validation studies.

About HighScope Educational Research Foundation

Founded in 1970, the HighScope Educational Research Foundation is an independent, nonprofit, 501(c) 3 organization with headquarters in Ypsilanti, Michigan. HighScope's mission is to change the trajectory of the world one child at a time. Its vision is widespread participatory education in which students and teachers are partners in shaping the learning experience. To this end, it engages in evaluative research, development of curriculum, training and assessment materials, and dissemination through educational services and publishing. These activities target teachers and service workers, primarily in early childhood programs. The HighScope Foundation also disseminates research findings to those who influence children's lives, such as teachers, service workers, parents, administrators, policymakers, academics, and researchers.

The HighScope Center for Early Education Evaluation (CEEE) is the research and evaluation arm of the HighScope Educational Research Foundation. Originally established as the research division of HighScope, the Center was created in July 2012 to better reflect its role as internal evaluator of HighScope's products and services, as well as external evaluator and consultant for states and provider agencies.

Inquiries about this report can be directed to:

Dr. Jeffrey A. Beal
Senior Research Associate, Center for Early Education Evaluation
HighScope Educational Research Foundation
600 North River Street
Ypsilanti, MI 48198
jbeal@highscope.org
734-485-2000, Ext. 253

COR for Kindergarten Overview

HighScope Educational Research Foundation developed a new assessment tool for use by kindergarten teachers. This tool, COR for Kindergarten, provides an observation-based authentic assessment that kindergarten teachers can successfully administer and reliably score. It offers kindergarten entry and ongoing information that teachers can use to inform instruction, families can use to make decisions about their children's learning opportunities, and other stakeholders can use to inform efforts to close the school readiness gap.

COR for Kindergarten is comprised of two assessments, the Kindergarten Entry Record (KER) and the Ongoing Record (OR), and four progress-monitoring tools. The KER consists of 20 items across the five domains identified by the National Education Goals panel as essential for later school success: Literacy and Language (LL), Mathematics (MATH), Health and Physical Development (HPD), Approaches to Learning (ATL), and Social and Emotional Development (SED) (Kagan, Moore, & Bredekamp, 1995). It is an observation-based assessment administered during the first few months of school. The KER measures what a child knows and can do at kindergarten entry.

The Ongoing Record (OR) is an observation-based assessment that measures the same five domains and consists of the 20 KER items and 10 more items that kindergarten teachers and experts identified as useful to track throughout the kindergarten year. Teachers administer the OR up to three times across the school year after completion of the KER. The OR measures attainment of end-of-year kindergarten standards.

We embedded progress-monitoring tools in COR for Kindergarten because teachers can use the assessment throughout the school year. The tools include letter identification charts for lowercase and uppercase letters, letter sound charts, and numeral identification charts. The charts pop up when the teacher scores the letter identification, letter sound, and numeral identification items, allowing the teacher to easily record the known letters, sounds, and numerals to monitor progress. Another progress-monitoring tool is the Progress Monitor Record (PMR). The PMR has all of the Common Core State Standards (CCSS) for language, literacy, and mathematics. We provide the PMR through the online system as a PDF. Teachers have the option to use the PMR when they score a child "Not Assessable at this time." For example, if a child scores "Not assessable at this time" for *Counting on from any number*, the teacher can choose to score the child's ability to count by tens and count by ones on the PMR. The teacher can use the PMR as often as necessary to monitor the progress of the child.

When teachers use both assessments and the progress-monitoring tools, they get a comprehensive view of what the child knows and is able to do at kindergarten entry and throughout the school year. The teachers can use the information they collect to inform instruction for individual students as well as small- and whole-group instruction. Teachers can detect learning problems early and quickly and provide interventions that they can use to monitor children's progress and determine if the intervention is working as expected. Knowing where each child is on the developmental progression, for key indicators in the domains crucial for future school success, is fundamental to understanding how to support each child's learning and development. COR for Kindergarten provides that information in one assessment system that teachers can successfully administer and reliably score.

Development of COR for Kindergarten

HighScope understands the burden that an observation-based assessment administered during the first few months of school places on kindergarten teachers. To make sure their assessment efforts are worthwhile, we recruited kindergarten teachers to assist in the development of the assessment from the very beginning. Practitioner participation helped ensure that the results provide useful and meaningful information and that the assessments are

- practical to administer,
- explicitly aligned to early childhood and kindergarten standards, and
- comprehensive enough to take the place of most locally required assessments they presently administer.

HighScope began developing COR for Kindergarten by creating an assessment blueprint, developing the items, and designing the assessment format the summer of 2014. During the 2014–2015 school year, five kindergarten teachers reviewed the assessment blueprint, the items, and the assessment format and made recommendations for improvement. The same teachers participated in cognitive labs to help revise the items, discard items that were too easy or too hard, and narrow the number of items in the item bank to those that they believed would provide meaningful information to inform instruction. Then, a larger group of teachers and administrators (24) participated in two focus groups designed to review the items still in the item bank for bias, age-appropriate language, and developmentally appropriate progressions. The focus group participants suggested changes in item wording and in the developmental progressions. They narrowed the number of items in the item bank further according to whether the item was likely to provide meaningful information that could be used to inform instruction. We subsequently revised the assessment system based on the group's recommendations.

Additionally, we asked eight subject matter experts to review the COR for Kindergarten to ensure that the assessment items accurately measure the constructs they were designed to measure and that they are fair and free of bias. We revised the assessment using suggested changes from the reviewers and used the subsequent draft during the pilot studies and the implementation studies. All of the expert reviewers have PhDs in their respective fields, have experience reviewing early childhood assessments, and have worked in major universities across the United States. The expert reviewers included author and consultant Sue Bredekamp, PhD; Deborah Stipek, PhD, Stanford University; P. David Pearson, PhD, UC Berkeley; Tanya Wright, Michigan State University; Doug Clements, PhD, University of Denver; Stephen Sanders, EdD University of South Florida; Fredrick Morrison, PhD, University of Michigan; and Nancy Eisenberg, PhD, Arizona State University.

Testing the COR for Kindergarten Assessments

HighScope conducted three sequential studies to test and refine the assessment items and data system: an initial pilot, a revised pilot, and an implementation study. We assessed 1,982 children across the studies. We performed psychometric analysis throughout the studies.

Initial Pilot Phase

HighScope conducted the pilot study in two phases. We designed the initial phase to test the usability and utility of the assessment and the assessment website's functionality and ease of use. During the 2015–2016 school year, 27 kindergarten teachers from 18 schools in four states participated in the initial phase of the pilot study. The schools consisted of 2 private and 22 state-funded Local Education Agencies and 3 Public School Academies that are geographically located in

rural, suburban, middle-size cities, urban fringe, and urban areas. The children came from a wide range of socioeconomic backgrounds based on the free- and reduced-lunch counts and submitted family surveys that asked about income. The schools administered the KER in September and October of 2015 and the OR in November and December of 2015 to 478 kindergarten children.

Initial Phase Results

Teachers completed surveys about their use of the website during and after each administration. We found that the assessment's website was easy to use with little training and teacher satisfaction with the website was extremely high. The KER we developed for the initial pilot consisted of 18 items across 5 domains. We used the data collected from teachers to determine whether each item was distinct, whether a 1-, 3-, or 5- dimensional model worked best, and finally whether each domain assessed a distinct developmental construct. We found that the 5-dimensional model worked best; that each of the 18 items were distinct but did not fit well together; and that there were too few items in some domains, resulting in extremely low reliabilities. We revised the KER by increasing the number of items to 30, extending the developmental progressions to five categories, and rewording items for developmental appropriateness in preparation for the second phase of the pilot. The OR we developed for the initial pilot phase consisted of 30 items across 5 domains, and we had similar findings for the OR that we had for the KER. We revised the OR by increasing the number of items to 40, extending the developmental progressions to five, and rewording items for developmental appropriateness in preparation for the second phase of the pilot.

Revised Pilot Phase

During the revised phase of the pilot, teachers rated 471 kindergarten children using the KER in February and March of 2016 and 467 kindergarten children using the OR in April and May of 2016. After the teachers completed each assessment, we examined whether the levels used to assess children's progress indeed follow the theorized developmental progression, and we conducted reliability and validity studies. We also examined whether the structure of the instrument is consistent in distinctly capturing the various developmental content areas. Dr. Everett Smith, Psychometric Consultant, analyzed the data on Conquest using the Multidimensional Partial Credit Rasch Model. The analyses identified 20 items for the KER that fit together and 25 items for the OR that fit together. The results follow.

Results for the Revised Pilot Phase for Kindergarten Entry Record

Student Sample: We describe the children who participated in the revised pilot phase study of the KER in Figure 1.

Figure 1

KER: Sample (n=471)

Gender:	
51% Girls	49% Boys
ELL:	
94% No	6% Yes
IEP:	
95% Yes	5% No
Free/Reduced Lunch:	
36% No	64% Yes
Race:	
60% White	
27% African American	
10% Hispanic	
2% More than one race	
1% Asian	

Scale Reliabilities (Chronbach's Alpha): We show the COR for Kindergarten Domain Scale reliabilities in Figure 2, along with the number of items per domain, for a sample of N=471 children attending kindergarten. Considering the number of measures were limited to 20 items across five domains in order to reduce the assessment burden on kindergarten teachers, the reliability coefficients are good. The Alphas are all above the acceptable threshold of .70.

Figure 2

KER: Scale Reliabilities

COR for Kindergarten Domains	# of Measures	Reliability
LL: Literacy and Language	6	.84
M: Mathematics	4	.82
HPD: Health and Physical Development	3	.78
ATL: Approaches to Learning	3	.86
SED: Social and Emotional Development	4	.91

(n=471)

Domain Correlations (Rasch): We show the intercorrelations among the domains of the COR for Kindergarten in Figure 3. Some of the scores are highly correlated and some are not, which appropriately reflects that young children's development is rapid, uneven, episodic, and highly influenced by the environment (Shepard, Kagan, and Wurtz, 1998).

Figure 3

KER: Domain Correlations

COR for Kindergarten Domains	Math	HPD	ATL	SED
LL	.845	.531	.831	.440
MATH		.452	.640	.383
HPD			.554	.809
ATL				.643

All correlations significant at *p*<.01

Model Fit, Wright Maps, and Item Curve Characteristics (Rasch): The weighted and unweighted Mean-Square fit statistics for the items were all below 1.4 with the exception of the gross motor item, which was marginal (weighted 1.49). We did not observe misfit for any item beyond that expected within the Partial Credit Rasch model with the exception of the gross motor item, which means that the items within each dimension are unidimensional or all measure the same construct. We revised the gross motor item for the implementation study.

The Wright Maps showed that our *a priori* predictions about the relative difficulty of items were correct. Items that are more difficult (i.e., retelling stories and writing) are at or near the top and items that are less difficult (i.e., letter identification and gross-motor skills), are at or near the bottom. A Wright Map of MLE Estimates and Thresholds showed the general orderliness with which the polytomous item steps group with developmental levels 1 at the bottom followed by levels 2 and levels 3, with levels 4 at the top, which means that the developmental progressions for each item are in the right order. The polytomous Item Curve Characteristics across the items provided further evidence of an orderly successive progression of developmental levels.

External Measures Correlations: A subset of students in the data collection (eight randomly chosen students from 26 of the participating classrooms) were directly assessed by HighScope research staff using valid and reliable instruments that measure the following domains: Literacy and Language, and Mathematics. Direct assessment allowed us to examine the relation between COR for Kindergarten KER and a battery of validated "gold standard" instruments that are often used in research studies (See Table 1). However, direct assessments and observation-based assessments do not assess the same behaviors in the same way; therefore, we did not expect the correlations to converge as though they do. We did expect the correlation values for the selected COR for Kindergarten domains to relate to these direct assessments positively, indicating they are measuring the same underlying constructs, and they did. We show the correlations between these "gold standard" measures and COR for Kindergarten Domains in Figure 4.

COR *for Kindergarten*

Table 1

List of External Measures Correlated with Corresponding COR for Kindergarten Domains

COR for Kindergarten External Measures Aligned to KER and Ongoing Record Items	
Instruments	COR for Kindergarten Domain
WJ-III Tests of Achievement	
Test 1 – Letter-Words Identification	Literacy and Language
Test 9 – Passage Comprehension	
Test 13 – Word Attack	
WJ-III Tests of Achievement	
Test 10 – Applied Problems	Mathematics
Test 18a and 18b – Quantitative Concepts	

Figure 4

Correlations between COR for Kindergarten KER & Other Measures

External Measures	COR for K LL	COR for K Math	COR for K ATL	COR for K SED
WJ III T1: Letter-Words Identification	0.70	0.48	0.44	0.30
WJ III T9: Passage Comprehension	0.66	0.42	0.45	0.29
WJ III T13: Word Attack	0.65	0.44	0.42	0.29
WJ III T10: Applied Problems	0.47	0.45	0.49	0.23
WJ III T18a/18b: Quantitative Concepts	0.54	0.48	0.55	0.33

All correlations significant at $p<.05$

Results for the Revised Pilot Phase — Ongoing Record

Student Sample: We describe the children who participated in the revised pilot phase study of the OR in Figure 5.

Figure 5

OR: Sample (n=467)

Gender:	
51% Girls	49% Boys
ELL:	
94% No	6% Yes
IEP:	
95% Yes	5% No
Free/Reduced Lunch:	
36% No	64% Yes
Race:	
59% White	
28% African American	
10% Hispanic	
2% More than one race	
1% Asian	

Scale Reliabilities (Chronbach's Alpha): We show the COR for Kindergarten OR Domain Scale reliabilities in Figure 6, along with the number of items per domain, for a sample of N=467 children attending kindergarten. There were 25 items on the OR including two additional Literacy and Language items and two additional Math items, which more than likely account for the increased scale reliabilities for those two domains.

Figure 6

OR: Scale Reliabilities

COR for Kindergarten Domains	# of Measures	Reliability
LL: Literacy and Language	8	.90
M: Mathematics	6	.89
HPD: Health and Physical Development	4	.78
ATL: Approaches to Learning	3	.85
SED: Social and Emotional Development	4	.91

(n=467)

Domain Correlations (RASCH): We show the intercorrelations among the domains of the COR for Kindergarten OR in figure 7. Again, some of the scores are highly correlated and some more moderately correlated, appropriately reflecting that young children progress at different rates in different areas of development.

Figure 7

OR: Domain Correlations

COR for Kindergarten Domains	Math	HPD	ATL	SED
LL	.798	.786	.760	.746
MATH		.678	.600	.604
HPD			.854	.864
ATL				.891

All correlations significant at *p*<.01

Model Fit, Wright Maps, and Item Curve Characteristics (RASCH): The weighted and unweighted Mean-Square fit statistics for the items were all below 1.4 meaning that we did not observe misfit for any item beyond that expected within the Partial Credit Rasch model. A Wright Map of MLE Estimates and Thresholds showed the general orderliness with which the polytomous item steps group with developmental levels. The polytomous Item Curve Characteristics across the items, however, showed disorder for three items: *Shows flexibility in shifting strategies, Reflects upon solutions,* and *Resolves conflicts.* We revised those three items for the implementation study.

Summary

Expected scale reliabilities and levels of intercorrelations between domains are evident in the data. With the exception of the gross motor item on the KER, the items fit well together. For the most part, the Wright Maps and the polytomous Item Curve Characteristics across the items provide evidence of an orderly successive progression of developmental levels with the exception of the three items on the OR that showed disorder. The correlations between the KER and external measures were positive. We revised the four items that were problematic and tested them in the implementation study.

Implementation Study

HighScope conducted an implementation study to investigate further the psychometric properties of COR for Kindergarten when used by a more diverse population of teachers and students. We examined the effects of revising the disordered gross motor item in the KER, revising the three disordered items in the OR, and adding five items, for a total of 30, to the OR assessment. During the 2016–2017 school year, 79 kindergarten teachers from 28 schools in seven states participated in the COR for Kindergarten implementation study. The schools consisted of 6 schools and 16 Local Education Agencies and 6 Public School Agencies that were geographically located in a mixture of rural, suburban, middle-size cities, urban fringe, and urban areas. Children came from a wide range of socioeconomic backgrounds based on their free/reduced lunch counts and a review of the cost of attending the private and unsubsidized schools. These teachers rated 1,506 kindergarten children using the KER and 738 kindergarten children using the OR.

After the teachers completed each assessment, we examined whether the levels used to assess children's progress indeed follow the developmental progression theorized and conducted reliability and validity studies. We also examined whether the structure of the instrument is consistent in distinctly capturing the various developmental content areas. Dr. Everett Smith,

Psychometric Consultant, analyzed the data on Conquest using the Multidimensional Partial Credit Rasch Model.

Implementation Study Results

Scale Reliabilities (Chronbach's Alpha): We show the COR for Kindergarten Domain Scale reliabilities in Figure 9, along with the number of items per domain, for a sample of N=1,506 children attending kindergarten. There was an increase from the pilot study in four domains; only the ATL domain remained the same.

Figure 9

KER: Scale Reliabilities

COR for Kindergarten Domains	# of Measures	Reliability
LL: Literacy and Language	6	.86
M: Mathematics	4	.87
HPD: Health and Physical Development	3	.86
ATL: Approaches to Learning	3	.86
SED: Social and Emotional Development	4	.93

(n=1,506)

Domain Correlations (Rasch): We show the intercorrelations among the domains of the COR for Kindergarten in Figure 10. Again, some of the scores are highly correlated and some are not, appropriately reflecting that young children progress at different rates in different areas of development.

Figure 10

KER: Domain Correlations

COR for Kindergarten Domains	Math	HPD	ATL	SED
LL	.720	.495	.601	.402
MATH		.362	.487	.345
HPD			.775	.877
ATL				.800

All correlations significant at $p<.01$

Model Fit, Wright Maps, and Item Curve Characteristics (Rasch): The weighted and unweighted Mean-Square fit statistics for the items were all below 1.4 meaning that we did not observe misfit for any item beyond that expected within the Partial Credit Rasch model. The Wright Maps and the polytomous Item Curve Characteristics across the items provide further evidence of an orderly successive progression of developmental levels.

Inter-rater Reliability: The effects of training on inter-rater agreement were determined through a study conducted by presenting trained teachers with 23 scenarios that described a child's behavior, such as, *Jeff was able to identify the following letters of the alphabet (A, J, e, f, B, l). What score would Jeff get for Letter Identification?*. The teachers used the COR for Kindergarten KER to rate each scenario. We used a proprietary survey tool, KeySurvey, to administer the inter-rater reliability assessment and to gather the data. We compared the teacher rater scores to the scores given by a team of HighScope raters who are familiar with using the instrument. Using the data collected from 24 teachers (raters) and 23 scenarios, we used three measurements of inter-rater agreement: exact and adjacent agreement percentages, Kappa for polytomous responses, and intra-class correlation using a two-way ANOVA random model. (Both scenarios and teachers are random samples of their respective populations.)

The average exact percentage among the 23 scenarios is about 85%, while the adjacent percentage is 95%. The average Kappa for multiple response options is 0.65; levels above .61 imply substantial amount of agreement. The ICC (reliability of the mean) is 0.993, (95% CI: 0.988, 0.996), implying a great degree of agreement among raters and suggesting that the scenarios, on average, were coded similarly among the 24 teacher raters.

Implementation Study Results: Ongoing Record

Scale Reliabilities (Chronbachis Alpha): We show the COR for Kindergarten OR Domain Scale reliabilities in Figure 11, along with the number of items per domain, for a sample of N=738 children attending kindergarten.

Figure 11

OR: Scale Reliabilities

COR for Kindergarten Domains	# of Measures	Reliability
LL: Literacy and Language	10	.90
M: Mathematics	8	.80
HPD: Health and Physical Development	4	.75
ATL: Approaches to Learning	4	.83
SED: Social and Emotional Development	4	.86

(n=738)

Domain Correlations (Rasch): We show the intercorrelations among the domains of the COR for Kindergarten OR in Figure 12. Again, some of the scores are highly correlated and some more moderately, appropriately reflecting that young children progress at different rates in different areas of development.

Figure 12

OR: Domain Correlations

COR for Kindergarten Domains	Math	HPD	ATL	SED
LL	.607	.683	.609	.508
MATH		.396	.616	.396
HPD			.641	.739
ATL				.750

All correlations significant at *p*<.01

Model Fit, Wright Maps, and Item Curve Characteristics (Rasch): The weighted and unweighted Mean-Square fit statistics for the items were all below 1.4, meaning that we did not observe misfit for any item beyond that expected within the Partial Credit Rasch model. The Wright Maps and the polytomous Item Curve Characteristics across the items provide further evidence of an orderly successive progression of developmental levels.

Summary

The implementation study results show that the COR for Kindergarten assessment provides reliable and valid psychometric measurement of the development of children entering kindergarten and throughout the kindergarten year for five key domains. Expected levels of intercorrelations between domains are evident in the data. The Wright Map shows that our *a priori* predictions about the relative difficulty of items were correct. Items that are more difficult (i.e., retelling stories and writing) are at or near the top and items that are less difficult (i.e., letter identification and gross-motor skills) are at or near the bottom. Our Wright Map of MLE Estimates and Thresholds shows the general orderliness with which the polytomous item steps group with developmental level 1 at the bottom, followed by levels 2 and levels 3, with level 4 at the top. The polytomous Item Curve Characteristics across the items provide evidence of an orderly successive progression of developmental levels.

Our challenge was to create a kindergarten entry assessment and ongoing assessment that conforms to early childhood assessment standards (observation-based authentic assessment), that is practical to administer during the first few months of kindergarten and throughout the year, and that provides useful and meaningful results teachers can use to inform their instruction. HighScope met that challenge in creating COR for Kindergarten.

References

Kagan, S.L., Moore, E., & Bredekamp, S. (Eds.). (1995, June). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Goal 1 Technical Planning Group Report 95–03. Washington, DC: National Education Goals Panel.

Shepard, L. A., Kagan, S. L., & Wurtz, E. (Eds.). (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Goals Panel.