K-3 FAP Research Update

UNC CHARLOTTE
College of Education

CEME
The Center for Educational Measurement and Evaluation

Statewide Kindergarten Formative Assessment Systems: Challenges and Innovative Designs

Reliability and Validity Evidences

Do-Hong Kim
Augusta University
Richard G. Lambert
University of North Carolina at Charlotte

Presentation to the Special Conference on Classroom Assessment, National Council for Measurement in Education, University of Kansas, September, 2017

The NC Kindergarten Entry Assessment

- The Office of Early Learning (OEL) at the North Carolina Department of Public Instruction implemented a new formative assessment process in all kindergarten classrooms in the state during the 2015-16 academic year.

- *The NC K-3 Formative Assessment Process: Kindergarten Entry Assessment* is the initial step in the development of a comprehensive formative assessment process for young children from kindergarten entry through the end of third grade

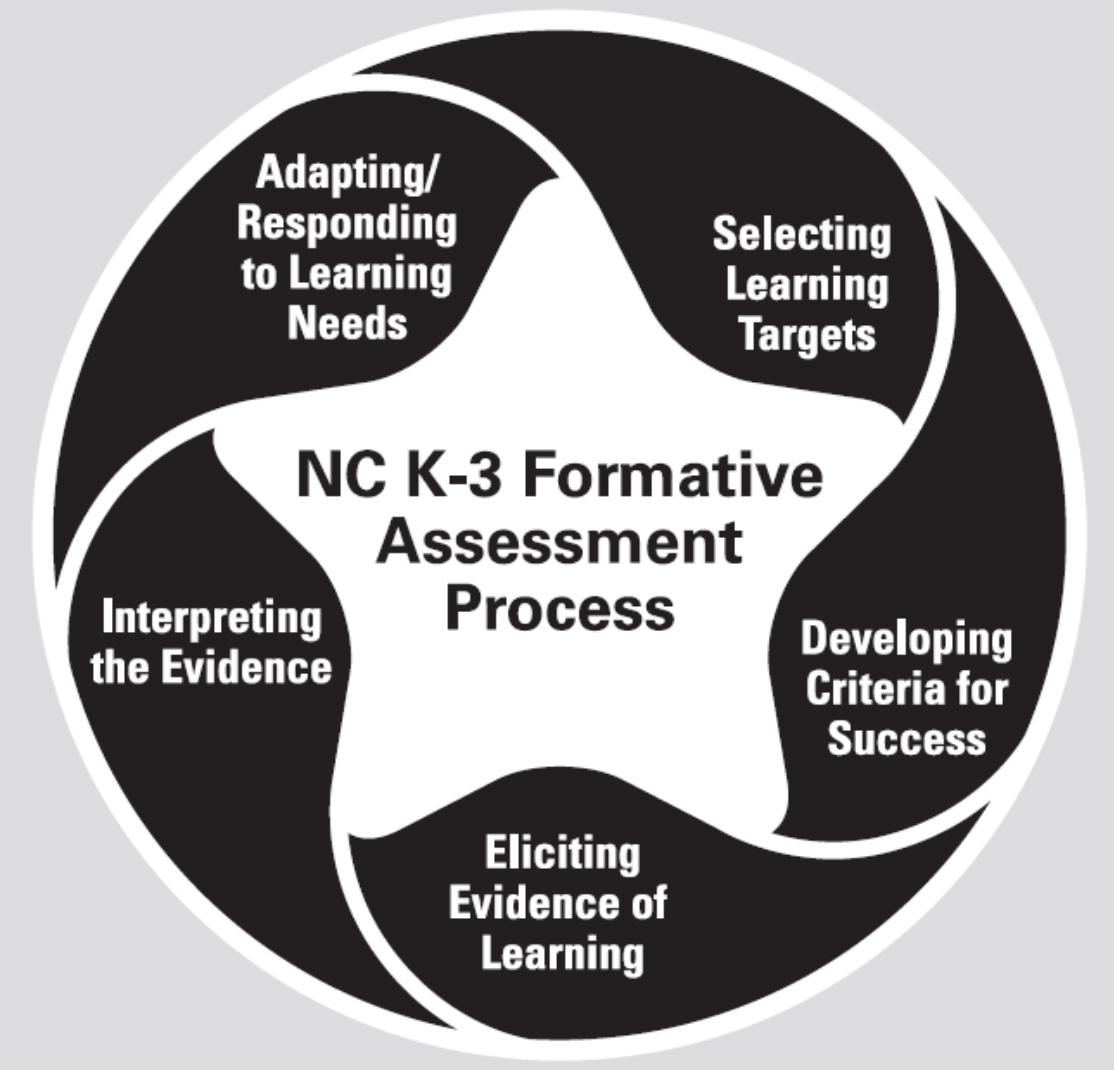
The NC Kindergarten Entry Assessment

- Completed within 60 days of enrollment
- Teachers gathered evidences of student learning, then uploaded those evidences to an online platform to monitor progress and inform their instruction
- Districts created their own implementation plans based on their unique capacities with guidance from state consultants assigned to each state board of education region

The NC Kindergarten Entry Assessment

- The assessment process includes five steps for teachers:
  - Selecting learning targets
  - Developing criteria for success
  - Eliciting evidence of student learning
  - Interpreting the evidence
  - Adapting instruction to respond to the student's learning needs.

NC Formative Assessment Process



The NC Kindergarten Entry Assessment

- To assist teachers in identifying and selecting learning targets for their students, the assessment includes several construct progressions that outline the developmental pathways of foundational skills within the five domains of early childhood development that are necessary for academic achievement.

- The kindergarten entry portion of the NC K-3 Formative Assessment Process (NC K-3 FAP) currently consists of 10 construct progressions.

Formative Assessment

- “By its nature and development, this process provides teachers with information to include and respond to individual student’s specific strengths as well as specific areas for growth. The NC K-3 Formative Assessment Process (FAP) is a part of daily instruction, rather than a formal one-on-one testing situation. Gathering information about student learning in a formative manner is a natural, ongoing process that is applicable to all students.”

- (NC Construct progressions and situations, Office of Early Learning, 2015)

Current Study

- To investigate the data collected across the whole state for the 8 required progressions in 2016, which was the second effective year of implementation.

Construct Progressions

Domain	Construct
Approaches to Learning	Engagement in Self-selected Activities
Emotional and Social Development	Emotional Literacy
Health and Physical Development	Grip and Manipulation
	Hand Dominance
	Crossing Midline
Cognitive Development	Object Counting
Language Development and Communication	Letter Naming
	Following Directions
	Book Orientation
	Print Awareness

KEA Electronic Portfolios

Teachers uploaded evidences such as notes, videos and pictures that were related to eight progressions:

- Engagement in Self-selected Activities (a scale of 0-5)
- Emotional Literacy (a scale of 0-10)
- Grip and Manipulation (a scale of 0-6)
- Hand Dominance (a scale of 0-5)
- Crossing Midline (a scale of 0-4)
- Object Counting (a scale of 0-9)
- Letter Naming (a scale of 0-12)
- Following Directions (a scale of 0-12)

Data

- All districts in the state participated in 2016:
  - ~1,300 elementary schools
  - ~6,000 kindergarten teachers
  - ~116,000 children

Data Analysis

- Classical Test Theory (CTT) analysis
  - Item difficulty and item discrimination
  - Reliability – Cronbach's alpha and Feldt-Gilmer congeneric reliability coefficients
  - The overall standard error of measurement using the traditional formula: $SEM=SD\sqrt{1\text{-reliability}}$
- Item Response Theory (IRT) analysis using Partial Credit Model (PCA) (Masters, 1982)
  - Dimension
  - Reliability
  - Item difficulty
  - Rating category effectiveness

CTT Item Difficulty

Construct (scale)	Item Difficulty	Item Discrimination
Engagement in Self-selected Activities (0-5)	2.50	0.81
Emotional Literacy (0-10)	4.15	0.81
Grip and Manipulation (0-6)	3.20	0.78
Hand Dominance (0-5)	1.16	0.32
Crossing Midline (0-4)	1.49	0.40
Object Counting (0-9)	3.91	0.82
Letter Naming (0-12)	6.84	0.81
Following Directions (0-12)	5.85	0.84

CTT Item Difficulty

Construct (scale)	Female	Male
Engagement in Self-selected Activities (0-5)	2.57	2.36
Emotional Literacy (0-10)	4.19	3.92
Grip and Manipulation (0-6)	3.27	3.00
Hand Dominance (0-5)	0.97	0.92
Crossing Midline (0-4)	1.39	1.33
Object Counting (0-9)	3.85	3.75
Letter Naming (0-12)	6.76	6.49
Following Directions (0-12)	5.87	5.49

CTT Item Difficulty

Construct (scale)	African American	Hispanic	White
Engagement in Self-selected Activities (0-5)	2.29	2.39	2.61
Emotional Literacy (0-10)	3.77	3.75	4.37
Grip and Manipulation (0-6)	3.04	3.11	3.22
Hand Dominance (0-5)	0.78	0.91	1.07
Crossing Midline (0-4)	1.42	1.37	1.32
Object Counting (0-9)	3.51	3.39	4.16
Letter Naming (0-12)	6.33	5.93	7.14
Following Directions (0-12)	5.30	5.23	6.14

CTT Item Discrimination

Construct (scale)	Female	Male
Engagement in Self-selected Activities (0-5)	0.83	0.81
Emotional Literacy (0-10)	0.82	0.81
Grip and Manipulation (0-6)	0.80	0.78
Hand Dominance (0-5)	0.26	0.27
Crossing Midline (0-4)	0.36	0.37
Object Counting (0-9)	0.82	0.82
Letter Naming (0-12)	0.82	0.80
Following Directions (0-12)	0.85	0.84

CTT Item Discrimination

Construct (scale)	African American	Hispanic	White
Engagement in Self-selected Activities (0-5)	0.80	0.82	0.83
Emotional Literacy (0-10)	0.80	0.81	0.82
Grip and Manipulation (0-6)	0.81	0.79	0.78
Hand Dominance (0-5)	0.24	0.25	0.27
Crossing Midline (0-4)	0.38	0.36	0.35
Object Counting (0-9)	0.81	0.81	0.82
Letter Naming (0-12)	0.80	0.78	0.82
Following Directions (0-12)	0.83	0.82	0.85

CTT Reliability

- Reliability was satisfactory as evidenced by
  - Cronbach alpha of .89 with SEM=5.27
  - Feldt-Gilmer estimate of .93 with SEM=4.21

- The reliability did not vary much across subgroups of the sample.

IRT Dimensionality

- 68.7% of variance accounted for by the measure.
- No strong evidence of the secondary dimension (the eigenvalue of the first contrast is small < 2.0)
- The mean square infit and outfit statistics for items (0.89 - 1.25) were well within acceptable limits

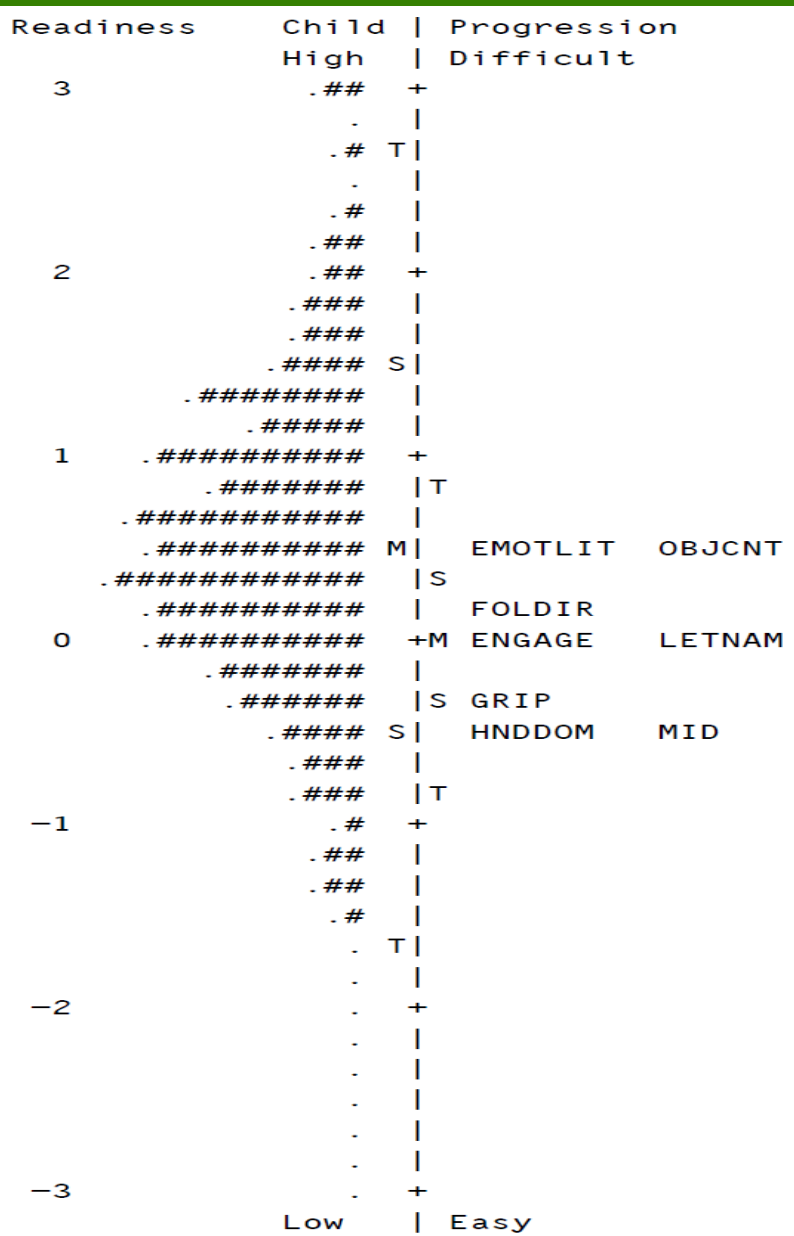
IRT Reliability

- Item reliability = .99
- Person reliability = .90
- Person Separation Index = 3.00
- Results provide strong evidence for the reliability of the measures.

IRT Item Difficulty

- The results of the PCM analysis showed that Emotional Literacy (*b*=0.57) was estimated to be the most difficult item, whereas Crossing Midline was estimated to be the easiest item (*b*= -0.52).

- The items had a similar relative difficulty across gender and ethnicity subgroups (DIF Contrast <0.5).


Readiness
Child
Progression
High
Difficult
3
2
1
0
−1
−2
−3
T
S
M
EMOTLIT
OBJCNT
FOLDIR
ENGAGE
LETNAM
GRIP
HNDDOM
MID
Low
Easy

IRT Item Difficulty

Construct	Total	Female	Male
Emotional Literacy	0.57	0.57	0.57
Object Counting	0.50	0.56	0.46
Following Directions	0.17	0.17	0.17
Engagement in Self-selected Activities	0.05	-0.08	0.10
Letter Naming	0.01	0.06	0.01
Grip and Manipulation	-0.27	-0.38	-0.20
Hand Dominance	-0.51	-0.59	-0.54
Crossing Midline	-0.52	-0.61	-0.48

IRT Item Difficulty

Construct	Total	African American	Hispanic	White
Emotional Literacy	0.57	0.57	0.57	0.57
Object Counting	0.50	0.53	0.50	0.48
Following Directions	0.17	0.17	0.17	0.17
Engagement in Self-selected Activities	0.05	0.14	0.05	-0.02
Letter Naming	0.01	-0.03	0.01	0.01
Grip and Manipulation	-0.27	-0.27	-0.27	-0.21
Hand Dominance	-0.51	-0.62	-0.51	-0.48
Crossing Midline	-0.52	-0.57	-0.52	-0.47

IRT Rating Category Effectiveness

- The observed means for all categories for all progressions were ordinal.
- Some disordered category probabilities were observed for Emotional Literacy, Hand Dominance, Object Counting, Letter Naming, and Following Directions.
- These issues appeared in the middle categories of the progressions, indicating issues with the discrimination of it by teachers, but the items still functioned well in terms of model fit.

Implementation Fidelity

- Each district conducted their own training
- Wide variability in training, supports, and fidelity
- Wide variability in quantity and quality of evidences
- Teacher face many challenges
- Time, Is the KEA a demand or a resource?
- Beliefs about formative assessment
- Misunderstanding about the purpose of the electronic portfolio
- Teachers who accessed the Live Binder collected twice as many evidences

The Live Binder

- Access the Live Binder and look over the resources and progressions.

- Discuss the evidences teachers could collect to support selection of learning status placements on the progressions.

http://www.livebinders.com/play/play/1606285#

Summary & Discussion

- These analyses are exploratory and diagnostic only
- The results provide encouraging evidence of reliability and validity of the KEA progressions
- Teachers may be need more training about the evidences that support the middle categories and may not be able to reliably distinguish between them
- Focused teacher training and professional development may be needed to help them more accurately assess students

Video interviews