What Works When Teaching Young Children to Read:

A Pilot Study of the Reliability of a Fidelity-of-Implementation Instrument

Atlanta, Georgia

April 8, 2005

FLORIDA INSTITUTE OF EDUCATION
at the University of North Florida

ELLM Presenter

Dr. Madelaine M. Cosgrove, Ed.D.

Associate Director for School Readiness, Florida Institute of Education PCER, Co-Investigator

The Preschool Curriculum Evaluation Research (PCER) program funded by the Institute of Education Sciences (IES), U.S. Department of Education includes a national evaluation study conducted by RTI International and Mathematica Policy Research (MPR), and complementary research studies conducted by each grantee. The findings reported here are based on the complementary research activities carried out by the Florida Institute of Education at the University of North Florida under the PCER program. These findings may differ from the results reported for the PCER national evaluation study. The findings presented in the Poster Symposium at the Society for Research in Child Development 2005, Biennial Meeting are based on a larger sample size of children, classroom and teachers and sought to answer complementary research questions including program effectiveness. The content of this presentation does not necessarily reflect the views or policies of the PCER Consortium including IES, RTI, and MPR, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Department of Education.

Introduction

This pilot study presents evidence of the reliability and validity of a fidelity-of-implementation instrument designed to measure teachers' instructional practices in using the critical components of emergent literacy identified by research as necessary to future reading and school success:

- Read Aloud to Children/Print Concepts
- Oral Language
- Emergent Writing
- Phonological Awareness
- Letter/Sound

The instrument was designed to measure two aspects of each component, *Pedagogical Quality* and *Time-on-Task*.

Methodology

To determine teachers' level of fidelity-of-implementation, teachers, asked to conduct lessons as usual, were videotaped for an entire day during March and again in May. Trained assessors segmented the tapes to (1) identify the implementation of the literacy components, (2) determine the total length in minutes of the segments using each literacy component (Time-on-Task), and (3) rate the levels of Pedagogical Quality using specific literacy-component items on the instrument. An observed item scores 1, an unobserved item scores 0. The component score is the proportion of the items observed.

Methodology 2

Scoring of Pedagogical Quality

Component	Number of Items
Print Concepts	17
Oral Language	19
Emergent Writing	12
Phonological Awareness	7
Letter/Sound	8

Methodology 3

Assessor Training

Assessors met weekly to practice segmenting the videotapes to identify literacy- component episodes, clarify component boundaries, and establish videotaping guidelines. Group and individual practice continued until assessors were consistently segmenting tapes, establishing *Time-on-Task*, and rating *Pedagogical Quality*. At this point, assessors were certified to use the fidelity-of-use instrument.

Methodology 4

Participants

Teachers participating in the PCER study were assigned to ELLM or wait-list control (W-L Control). ELLM teachers received intensive 2-day summer training and ongoing professional development support focused on helping them to become more proficient in implementing instructional strategies and activities that address the five components. W-L Control teachers received no training in addition to that routinely provided by the curriculum developers and center directors.

Methodology 5

Table of Participating Teachers

Time	Number of ELLM Teachers	Number of W-L Control Teachers
Time 1	17	10
Time 2	19	10
Both Times	12	5

Results

The pilot study of the fidelity-of-use instrument presents evidence of the reliability and validity of scores obtained by trained assessors using the instrument.

Pedagogical Quality, Mean Percent of Items, Time 1

Trait	W L Control (n=10)	ELLM (n=19)
Print Concepts	14.7	63.5
Oral Language	75.3	75.7
Emergent Writing	25.0	44.7
Phonological Awareness	6.3	50.7
Letter/Sound	38.6	51.1

Pedagogical Quality, Mean Percent of Items, Time 2

Trait	W L Control (n=10)	ELLM (n=20)
Print Concepts	52.4	62.9
Oral Language	76.3	75.0
Emergent Writing	32.5	35.8
Phonological Awareness	7.5	29.4
Letter/Sound	24.3	55.0

Results 2

The pilot study of the fidelity-of-use instrument presents evidence of the reliability and validity of scores obtained by trained assessors using the instrument.

Time on Task, Mean Time on Task (in minutes), Time 1

Trait	W L Control (n=10)	ELLM (n=19)
Print Concepts Time	4.4	8.4
Oral Language Time	12.1	11.0
Emergent Writing Time	4.6	13.7
Phonological Awareness Time	0.4	5.6
Letter/Sound Time	2.9	11.3

Time on Task, Mean Time on Task (in minutes), Time 2

Trait	W L Control (n=10)	ELLM (n=20)
Print Concepts	8.9	13.3
Oral Language	27.1	11.4
Emergent Writing	6.5	13.6
Phonological Awareness	1.0	1.9
Letter/Sound	2.9	6.4

Results 3

Reliability

Cronbach's coefficient alpha was calculated for each *Pedagogical Quality* component for each videotaping time. The results were averaged to form overall measures of internal consistency.

Results 4

Cronbach's Coefficient Alpha of Internal Consistency

Category	Number 1 (n=29)	Number 2 (n=30)	Average
Print Concepts	.9365	.9390	.9378
Oral Language	.7019	.6902	.6961
Emergent Writing	.8379	.8952	.8666
Phonological Awareness	.9006	.9209	.9108
Letter/Sound	.9105	.9353	.9229

The internal consistency for the *Pedagogical Quality* component items was consistent across times and ranged from a low of .69 on Oral Language to a high of .94 on Print Concepts. Inner-rater reliability for *Time-on-Task* and *Pedagogical Quality* is yet to be determined.

Results 5

Validity

Evidence supporting the validity of the *Pedagogical Quality* scores comes from three sources:

- External construct validity using a Multitrait-Multimethod Matrix to present evidence of trait validity.
- Construct validity evidenced by the convergence of the measures across groups. Because videotaped teachers are either ELLM or W-L Control, their Pedagogical Quality scores should differ in predictable ways.
- Evidence of nomological validity showing that the theoretical basis of the Pedagogical Quality scores of teachers provides a link to their students' literacy-related outcomes.

Results 6

Trait Validity: Multitrait-Multimethod

The coefficients in the validity diagonal (bold lower left) provide convergent evidence. The lack of significant correlations in the yellow triangles provides evidence that measures discriminate among traits.

Results 7

Multitrait-Multimethod Matrix

Pedagogical Quality - Pedagogical Quality

Trait	PC	OL	EW	PA	L/S
Print Concepts (PC)	**.94**	-	-	-	-
Oral Language (OL)	.40	**.70**	-	-	-
Emergent Writing (EW)	.00	-.20	**.87**	-	-
Phonological Awareness (PA)	*.41*	-.24	-.04	**.91**	-
Letter/Sound (L/S)	.33	.23	.23	.28	**.92**

Pedagogical Quality - Time on Task

Trait	PC	OL	EW	PA	L/S
Print Concepts (PC)	**-**	-	-	-	-
Oral Language (OL)	-	**-**	-	-	-
Emergent Writing (EW)	-	-	**-**	-	-
Phonological Awareness (PA)	-	-	-	**-**	-
Letter/Sound (L/S)	-	-	-	-	**-**

Time on Task - Pedagogical Quality

Trait	PC	OL	EW	PA	L/S
Print Concepts (PC)	***.79***	*.40*	-.20	.07	.39
Oral Language (OL)	-.17	**.51**	-.40[1]	-.46	-.05
Emergent Writing (EW)	.11	-.27	**.69**	.26	-.10
Phonological Awareness (PA)	.04	*-.47*	.09	**.63**	.26
Letter/Sound (L/S)	.44	.15	.35	.08	**.70**

Time on Task - Time on Task

Trait	PC	OL	EW	PA	L/S
Print Concepts (PC)	*	-	-	-	-
Oral Language (OL)	.01	*	-	-	-
Emergent Writing (EW)	-.15	-.29	*	-	-
Phonological Awareness (PA)	-.05	-.38[1]	.26	*	-
Letter/Sound (L/S)	*.52*	.12	.21	.08	*

n=22 time1 and time2 averaged scores.
Numbers on the main diagonal are reliabilities.
*Represents the non-available reliabilities of the Time on Task scores.
Bold-faced and italicized correlations are significantly different from zero, a=.05.

[1] Indicates a=.10
Lower diagonal bold-faced numbers represent monotrait-heteromethod correlations.
Numbers in the green triangles represent heterotrait-monomethod correlations.
Numbers in the yellow triangles represent heterotrait-heteromethod correlations.

Results 8

Trait Validity: Multitrait-Multimethod

The significant correlations between the scores of Oral Language and other components indicate Oral Language items do not distinguish the component from the other literacy-related components. This, together with the consistent *Pedagogical Quality* mean scores across ELLM and W-L Control teachers and the lowest of the reliabilities, led to the decision to remove this component from the pilot study.

Results 9

Multitrait-Multimethod Matrix

Pedagogical Quality - Pedagogical Quality

Trait	PC	EW	PA	L/S
Print Concepts (PC)	**.94**	-	-	-
Emergent Writing (EW)	.00	**.87**	-	-
Phonological Awareness (PA)	*.41*	-.04	**.91**	-
Letter/Sound (L/S)	.33	.23	.28	**.92**

Pedagogical Quality - Time on Task

Trait	PC	EW	PA	L/S
Print Concepts (PC)	**-**	-	-	-
Emergent Writing (EW)	-	**-**	-	-
Phonological Awareness (PA)	-	-	**-**	-
Letter/Sound (L/S)	-	-	-	**-**

Time on Task - Pedagogical Quality

Trait	PC	EW	PA	L/S
Print Concepts (PC)	***.79***	-.20	.07	.39
Emergent Writing (EW)	.11	**.69**	.26	-.10
Phonological Awareness (PA)	.04	.09	**.63**	.26
Letter/Sound (L/S)	.44	.35	.08	**.70**

Time on Task - Time on Task

Trait	PC	EW	PA	L/S
Print Concepts (PC)	*	-	-	-
Emergent Writing (EW)	-.15	*	-	-
Phonological Awareness (PA)	-.05	.26	*	-
Letter/Sound (L/S)	*.52*	.21	.08	*

Results 10

Construct Validity: Convergence Across Groups

ELLM teachers have experienced targeted professional development designed to improve critical literacy-related skills; therefore, higher Pedagogical Quality scores were expected.

ANOVA Table

Pedagogical Quality	Source of Variance	F Ratio Time 1 (n=29)	Probability Time 1	F Ratio Time 2 (n=30)	Probability Time 2
Print Concepts	Status*Education	0.02	.8857ǂ	7.83	.0095ǂ
Print Concepts	Education	0.12	.7304	-	-
Print Concepts	Status	29.24	<.0001	-	-
Emergent Writing	Status*Education	0.20	.6602ǂ	0.64	.4317ǂ
Emergent Writing	Education	0.01	.9045	0.58	.4522
Emergent Writing	Status	4.38	.0458	0.08	.7832
Phonological Awareness	Status*Education	0.55	.4664ǂ	0.17	.6879ǂ
Phonological Awareness	Education	0.01	.9045	0.47	.4970
Phonological Awareness	Status	15.78	.0458	3.12	.0880**
Letter/Sound	Status*Education	2.57	.1212ǂ	0.00	.9576ǂ
Letter/Sound	Education	0.60	.4447	1.71	.2020
Letter/Sound	Status	0.70	.4095	3.95	.0568**

Note: ǂ Indicates an education level and status interaction that was not significant and was removed from the analysis.
* Indicates a significant effect at a = .05. ** Indicates a significant effect at a=.10

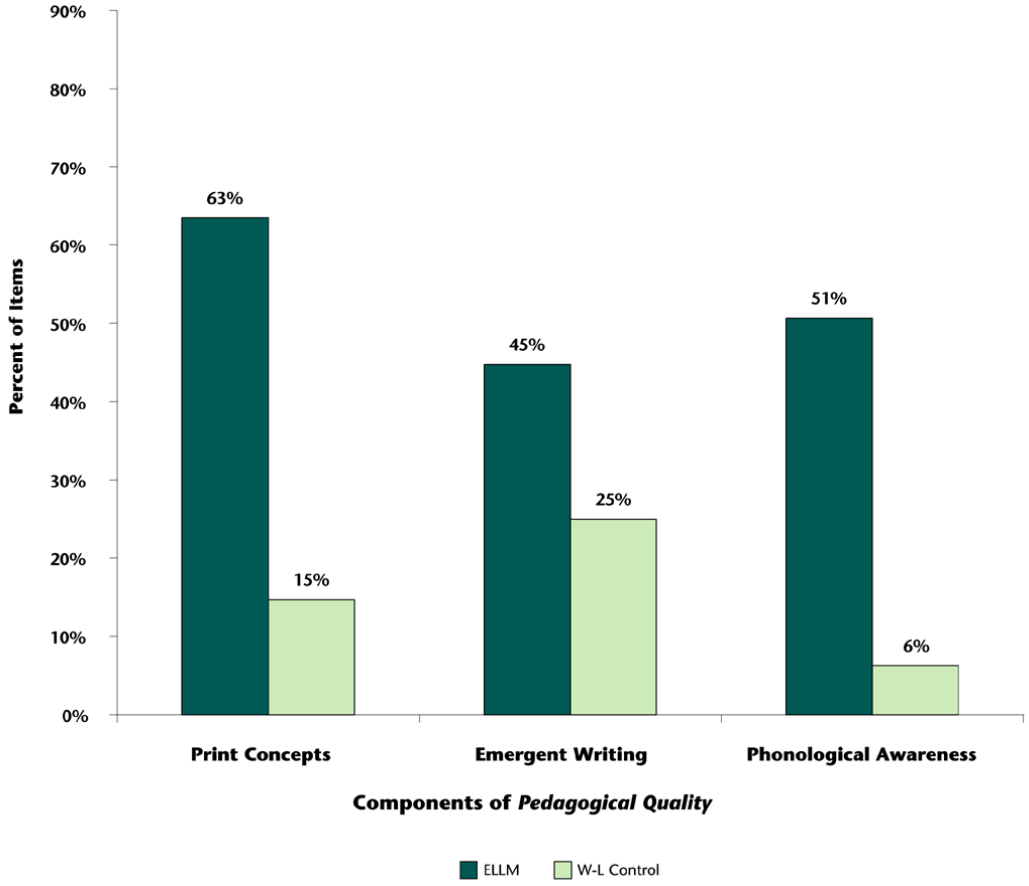
Results 11

Construct Validity: Convergence Across Groups

At Time 2, ELLM teachers, regardless of educational level, have higher *Pedagogical Quality* scores for Phonological Awareness and Letter/Sound. Additionally, at Time 2, ELLM teachers have higher *Pedagogical Quality* scores for Print Concepts than W-L Control non-degreed teachers.

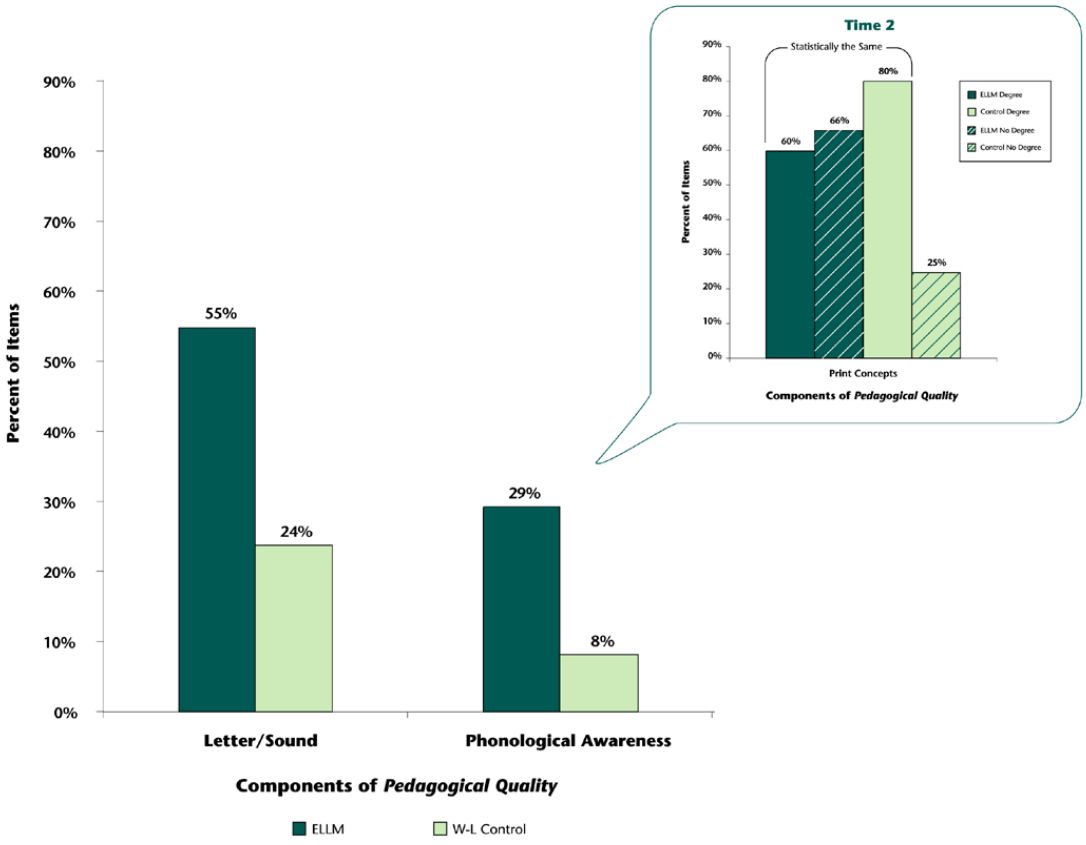
Results 12

Time 1: March 2003



Results 13

Time 2: May 2003



Results 14

Nomological Validity: Links to Student Outcomes

Teachers' scores on the Pedagogical Quality measure of the components were expected to correlate with students' literacy-related outcomes, as measured by the Test of Early Reading Ability, Third Edition (TERA-3) and the Alphabet Letter Recognition Inventory (ALRI). The ALRI measures a child's ability to recognize all upper- and lowercase letters of the alphabet arranged in non-alphabetical order.

Results 15

First-Order Correlations

Test	Posttest	Pedagogical Quality Component Correlations
TERA-3	Reading Quotient	Print Concepts (.1787)
TERA-3	Conventions of Print	Print Concepts (.1544), Letter/Sound (.1184)
TERA-3	Meaning	Print Concepts (.2440), Phonological Awareness (.1668)
ALRI	_	Emergent Writing (.1993)

Note: Numbers in parenthesis are correlations.

With the exception of the correlations involving the TERA-3 Meaning subtest, the correlations converged as expected. The *Pedagogical Quality* of Print Concepts is measured during segments in which the teacher is reading aloud to children; therefore, it is not surprising the children also gain understanding that print conveys meaning during these activities.

Conclusion

Evidence of the reliability of *Pedagogical Quality* was found using the Cronbach's coefficient alpha measure of internal consistency. Evidences of three types of validity were also presented for *Pedagogical Quality*.

With the exception of Oral Language, evidence of trait validity was presented for *Pedagogical Quality* for each of the critical literacy components. Analyses of the videotapes indicate that ELLM teachers' *Pedagogical Quality* scores on the literacy components were generally higher than the scores of the W-L Control teachers.

These differences in *Pedagogical Quality* scores were found regardless of education level, with the exception of one comparison. Furthermore, the *Pedagogical Quality* scores on the literacy components are positively correlated with students'posttest scores on the TERA-3 and ALRI.

Next Steps

- Provide evidence of inter-rater reliability for both the Time-on-Task and Pedagogical Quality measures of teachers' use of the critical literacy components in their instruction.
- Restructure the Oral Language component items to clearly establish distinguishable boundaries for the construct.
- Improve the Letter/Sound measure. The teachers' scores should correlate with students' test scores of alphabet letter knowledge.
- Provide evidence of the validity of the Time-on-Task measures.
- Use the videotape fidelity-of-use instrument to develop a classroom observation scale to measure teachers' use of the critical components of emergent literacy.

For more information visit us at:

www.unf.edu/dept/fie/ellm