

**Evaluation of the Core Knowledge
Program in Arkansas**

**Robert H. Bradley, Ph.D.
Evaluator**

November 28, 2005

Introduction

This report describes an evaluation of the Core Knowledge (CK) curriculum as implemented in Southeast Arkansas. CK is a full-spectrum academic program for 3- and 4-year old children. The program is designed to prepare the children for school success. Teachers are carefully trained in how to use program materials so as to enhance children's competence in basic academic skills.

The questions addressed in this evaluation are:

1. Do children who participated in CK show significant academic gains during the year?
2. Do children who participated in CK show significant gains in social competence during the year?
3. Do children with prior experience in CK continue to show significant academic progress during kindergarten and first grade?
4. Is the quality of teaching related to student academic gains during preschool?

Study Procedures

To answer Questions #1 and #2, children who participated in Core Knowledge programs during the 2004-05 school year were assessed during the late Fall of 2004 and the late Spring of 2005 using the Woodcock-Johnson – III achievement battery, the Phonological Awareness Literacy Scale –PK, and the Social Skills Raging System. Student performance in the Spring was compared to their performance in the Fall to determine whether there was significant improvement in academic skills.

To answer Question #3, children who participated in Core Knowledge programs during the 2002-03 and the 2003-04 school years were assessed during the late Fall of 2004 and the late Spring of 2005 using the Woodcock-Johnson – III achievement battery, the Phonological Awareness Literacy Scale, and the Social Skills Raging System. Student performance in the Spring was compared to their performance in the Fall to determine whether there was significant improvement in academic skills.

To answer Question #4, Core Knowledge teachers were assessed with the Early Language and Literacy Classroom Observation toolkit during the Fall of 2004 and again during the Spring of 2004. Children’s spring scores on Woodcock-Johnson, SSRS and PALS were regressed on their Fall scores plus teacher ELLCO scores to determine the extent to which Spring performance (or more specifically changes in performance from Fall to Spring) were a function of differences in teaching quality.

Child Assessments

The **Woodcock-Johnson III Tests of Achievement** battery was administered to participating children during the late Fall of 2004 and again during the late Spring of 2005. The WJ-III is a nationally-normed achievement battery of standardized assessments covering a wide range of academic subject matters. For this evaluation, 9 tests from the WJ-III battery were administered. Care was taken to select those tests most appropriate for evaluating the CK curriculum.

1. Letter-Word Identification. This test measures the student's word identification skills. Items for young children primarily involve identifying specific letters of the alphabet.
2. Story Recall. This test measures aspects of oral language including language development and meaningful memory. The tasks require the student to recall increasingly complex stories that are presented using an audio recording. After listening to a passage, the individual is asked to recall as many details of the story as possible.
3. Understanding Directions. This test requires a student to listen to a sequence of audio-recorded instructions then follow the directions by pointing to various objects in a colored picture. Items gradually increase in linguistically complexity.
4. Spelling. This test measures the student's ability to write orally presented words correctly. However, for young children, the focus is on prewriting skills such as drawing lines and tracing letters.
5. Listening Comprehension. This test involves symbolic learning, or the ability to match a pictographic representation of a word with the actual picture of the object.
6. Picture Vocabulary. This measure assesses oral language development and lexical (word) knowledge. The task requires the individual to identify pictured objects. Although a few receptive items are included at the beginning of the test, this is primarily an expressive language measure focused on single words.
7. Oral Comprehension. This test focuses on the student's ability to comprehend short audio-recorded passages and then supply the missing word using syntactic and semantic cues.
8. Quantitative Concepts. This test measures knowledge of mathematical concepts, symbols, and vocabulary. The test consists of two subtests. In the first subtest, the initial items require counting and identifying numbers, shapes, and sequences. Young children do not take the second subtest.
9. Applied Problems. This test requires the student to analyze and solve math problems. To solve the problems, the individual must listen to the problem, recognize the procedure to be followed, and then perform relatively simple calculations.
10. Academic Knowledge. This test includes three subtests: Science, Social Studies, and Humanities. It samples the student's knowledge of the sciences, history, geography, government, economics, art, music, and literature. Then early items require only a pointing response.

Scores on these 10 tests not only provide information on children's knowledge in the particularly areas examined but can be combined to form five more general cluster scores as well. The Oral Language – Standard cluster score is formed from scores on Story Recall and Understanding Directions. It is an aggregate measure of linguistic competency, listening ability, and

comprehension. The Oral Language – Extended cluster score is derived from a combination of Story Recall, Understanding Directions, Picture Vocabulary, and Oral Comprehension. It is an aggregate measure of expressive vocabulary, reasoning, listening comprehension, and memory. Listening Comprehension is an aggregate measure of listening ability and verbal comprehension. It is derived from scores on Understanding Directions and Oral Comprehension. Oral Expression is a combination of Story Recall and Picture Vocabulary. It is an aggregate measure of linguistic competency and expressive vocabulary. Math Reasoning is derived from Applied Problems and Quantitative Concepts. It provides a measure of mathematical knowledge and reasoning.

The **Social Skills Rating System** was also administered to children in the CK program during Fall of 2003 and Spring of 2004. It was also given Fall and Spring to CK graduates during the Kindergarten year and their matched controls. The SSRS is a nationally-normed measure of classroom behavior, focused on student-teacher relationships, peer acceptance, and academic performance. Specifically, it assesses the following areas.

1. Cooperation. Helping others, sharing materials, complying with rules and regulations.
2. Assertion. Initiating behaviors, asking others for information, introducing oneself, responding to actions of others.
3. Self-Control. Responding appropriately to teasing, taking turns, compromising.
4. Externalizing Problems. Aggression, poor control of temper, arguing.
5. Internalizing Problems. Anxiety, sadness, loneliness, poor self-esteem.
6. Hyperactivity. Excessive movement, fidgety, impulsive reactions. The Hyperactivity items are not given to preschool age children.

Items in the first three areas are combined to produce an overall Social Skills score and items on the final three are combined to produce an overall Problem Behaviors score.

The **Phonological Awareness Literacy Screening for Preschool (PALS-PreK)** was also administered to participating children during the late Fall of 2003 and late Spring of 2004 as part of the overall battery of achievement measures. Phonological awareness is the ability to identify, and reflect on various speech sounds. PALS-PreK is a standardized criterion-referenced test that assesses rhyme and beginning sounds. Specifically, it assesses: (1) Rhyme Awareness, (2) Alphabet Knowledge, (3) Beginning Sound Awareness, (4) Letter Sounds, (5) Print Knowledge, and (6) Name Writing.

Teacher Assessments

The **Early Language and Literacy Classroom Observation (ELLCO)** toolkit was administered to Core Knowledge teachers in the Fall of 2003 and the Spring of 2004. ELLCO assesses the extent to which teachers implement the CK program in accordance with program principles and objectives. As such, it is a measure of teacher quality. Specifically, through a combination of observation and interview, teachers are evaluated on: (1) the classroom literacy environment, (2) the general classroom environment, (3) the quality of language and literacy curriculum, and (4) the quality of specific literacy activities.

Study Design

To answer the four questions posed for this evaluation, a “pretest-posttest” design was used. This design was used because a true experimental design (i.e., one involving random assignment of 3- and 4-years olds to CK, other formal preschool, and no formal preschool groups) was neither ethically, politically, or fiscally feasible. The limitations of non-experimental designs are well-known. Accordingly, a variety of measurement and analytic strategies were used so as to improve confidence in the results obtained. Despite these efforts to improve the interpretability of findings, it is important to remember that non-experimental and quasi-experimental designs still have limitations. Thus, *any conclusions drawn from this study should be made with caution*, knowing both the potential strengths and the potential limitations of the pretest-posttest design.

Results

Prior to analyzing the data, all scores in the data set were reviewed to determine if values entered into the data set were out-of-range for a particular measure or were otherwise suspicious. No such values were found.

The primary statistical technique used to analyze data on student achievement was the dependent t-test. That procedure is appropriate when there are known dependencies in the data such as would occur since the same students took both pre- and post tests. Scores on the component scales from PALS-PreK, WJ-III, and SSRS were used as dependent variables in these analyses.

Regression analysis was used to analyze data pertaining to the effects of teaching quality on student outcomes (Question #4). Specifically, scores on PALS, WJ-III, and SSRS for the Spring were regressed on ELLCO scores and the comparable score on the student measures for the Fall. For example, the Spring score on Letter-Word Identification was regressed on ELLCO score plus the Fall score on Letter-Word Identification.

Woodcock-Johnson Achievement Test

During the Fall of 2004 and the Spring of 2005, 251 children currently in the CK program were assessed using several tests from the WJ-III achievement battery. Children who attended CK during the 2002-2003 and 2003-2004 academic years were also assessed in the Fall of 2004 and Spring of 2005 (beginning and end of their school year) with the WJ-III.

In the Spring of 2005 complete WJ-III scores were available on 63 students who participated in CK during the 2002-2003 academic year and 86 students who participated in CK during the 2003-2004 academic year. These students attended a variety of area public schools. Data from these children were used for purposes of the evaluation. Raw scores on each test were converted to two types of norm-referenced scores: W-scores and standard scores. Both are quite useful for purposes of statistical analysis. W-scores are a special transformation of the Rasch ability scale. The W scale for each test is centered on a value of 500, which has been set to the approximate average performance of a 10-year old. Students whose performance is lower than the average score of a 10-year old receive scores below 500. Those with scores higher than the average score of a 10-year old receive scores greater than 500. Because tests on the WJ-III tap such a wide range of ability in each competence area, scores vary greatly; and the use of the Rasch scale allows evaluators to record changes in actual ability within or across years. By comparison, the standard score constantly corrects for average improvements that come with age. At each age level, scores are recalibrated so that the mean is set to 100 and the standard deviation to 15. Like IQ the average score on each WJ-III test at each age is 100. For purposes of this evaluation, both W-scores and standard scores were analyzed.

Table 1
**Performance of 2004-05 Core Knowledge Students on the Woodcock-Johnson III
 Achievement Test Battery – Pretest vs. Posttest**

Achievement Test	<u>Mean W-Scores</u>			<u>Mean Standard Scores</u>		
	CK	Comparison	Sig.	CK	Comparison	Sig.
<i>Individual Tests</i>						
Letter-Word Identification	329	352	<.01	95	101	<.01
Story Recall	469	475	<.05	100	104	<.01
Understands Directions	432	457	<.01	100	104	<.01
Spelling	367	389	<.01	90	96	<.01
Listening Comprehension	447	455	<.10	97	99	NS
Picture Vocabulary	452	460	<.01	97	100	<.01
Oral Comprehension	434	447	<.01	93	97	<.01
Quantitative Concepts	414	421	<.01	89	93	<.01
Applied Problems	407	412	<.05	93	96	<.01
Academic Knowledge	435	446	<.01	92	95	<.01
<i>Cluster Scores</i>						
Oral Language Standard	459	466	<.01	97	100	<.01
Oral Expression	470	476	<.01	97	100	<.01
Math Reasoning	395	411	<.01	85	92	<.01

As Table 1 shows, mean scores on some individual tests and cluster scores from the Woodcock-Johnson were similar for current CK students were approximately the same as the mean (100) for the national norms. More to the point, however, CK student showed important gains in every area tested during the school year. The gains in W-scores are not remarkable in the sense that children should make absolute improvements in each skill area as a function of both maturation and attendance in preschool. What is perhaps more important is the significant improvements in standard scores in most areas. This means that CK students actually made more improvements (on average) than members of the norm group.

Table 2
Performance of 2003-04 Core Knowledge Students on the Woodcock-Johnson III Achievement Test Battery during Kindergarten – Pretest vs Posttest

Achievement Test	<u>Mean W-Scores</u>			<u>Mean Standard Scores</u>		
	CK	Comparison	Sig.	CK	Comparison	Sig.
<i>Individual Tests</i>						
Letter-Word Identification	439	449	<.01	93	94	NS
Story Recall	363	400	<.01	99	106	<.01
Understands Directions	482	489	<.01	104	107	<.10
Spelling	454	463	<.01	104	105	NS
Listening Comprehension	460	468	<.01	100	102	<.01
Picture Vocabulary	411	427	<.01	100	99	NS
Oral Comprehension	419	438	<.01	94	101	<.01
Quantitative Concepts	423	440	<.01	102	111	<.01
Applied Problems	402	429	<.01	99	107	<.01
Academic Knowledge	422	443	<.01	90	99	<.01
<i>Cluster Scores</i>						
Oral Language Standard	468	474	<.01	99	102	<.01
Oral Expression	477	482	<.01	99	101	<.05
Math Reasoning	377	415	<.01	99	108	<.01

As Table 2 shows, students who attended CK continued to show academic improvements during the kindergarten year. As expected, they made absolute improvements in skills (i.e., significant changes in W-scores) in every area tested. More importantly, they continued to show more improvement than members of the norm group in most areas, including all three cluster scores; that is, their standard scores showed significant change during the year.

Table 3

Performance of 2002-03 Core Knowledge Students on the Woodcock-Johnson III Achievement Test Battery during First Grade – Pretest vs. Posttest

Achievement Test	<u>Mean W-Scores</u>			<u>Mean Standard Scores</u>		
	CK	Comparison	Sig.	CK	Comparison	Sig.
<i>Individual Tests</i>						
Letter-Word Identification	453	458	NS	91	93	NS
Story Recall	411	440	<.01	101	105	<.05
Understands Directions	488	492	<.05	105	106	NS
Spelling	466	471	<.05	101	102	NS
Listening Comprehension	471	476	<.01	100	101	NS
Picture Vocabulary	439	458	<.01	95	99	<.01
Oral Comprehension	447	458	<.01	100	103	<.05
Quantitative Concepts	423	440	<.01	102	111	<.01
Applied Problems	436	454	<.01	102	106	<.01
Academic Knowledge	451	459	<.01	98	99	NS
<i>Cluster Scores</i>						
Oral Language Standard	477	481	<.01	100	101	NS
Oral Expression	483	486	<.01	100	101	NS
Math Reasoning	427	448	<.01	102	105	<.05

As Table 3 shows, CK students continued to make significant academic progress during first grade. They showed significant absolute gains in every area tested but one (Letter-Word Identification). They also showed continued improvement relative to the norm group in about half the areas tested. This latter result strongly suggests that children's experiences in Core Knowledge in the preschool years established a base of competence that allowed them to continue to benefit from later schooling experiences.

Table 4
Performance of Students Who Participated in the Core Knowledge Program during 2004-2005 on the Woodcock-Johnson III Achievement Test Battery in the Spring of the Preschool Year (2005) as a Function Teaching Quality (ELLCO scores), Controlling for Achievement Test Performance in the Fall of 2004

Achievement Test	Quality of Teaching Effect	Significance Level
<i>Individual Tests</i>		
Letter-Word Identification	Yes	<.01
Story Recall	Yes	<.01
Understands Directions	Yes	<.01
Spelling	Yes	<.01
Listening Comprehension	Yes	<.01
Applied Problems	No	NS
Picture Vocabulary	Yes	<.01
Oral Comprehension	Yes	<.01
Quantitative Concepts	Yes	<.05
Academic Knowledge	Yes	<.01
<i>Cluster Scores</i>		
Oral Language Standard	Yes	<.01
Oral Expression	Yes	<.01
Math Reasoning	Yes	<.01

As Table 4 shows, there were significant differences on most Woodcock-Johnson tests at the end of the preschool year for CK students depending on the quality of teaching observed in the CK classroom. Teachers who were observed to be more fully and faithfully implementing the CK curriculum tended to have students whose WJ-III scores showed substantial impact during the academic year. This contrasts to teachers whose ELLCO scores were lower. Their gains were less evident.

Phonological Awareness Literacy Screening Test

Complete PALS-PreK scores were available on 229 CK students in both Fall and Spring.

The dependent t-tests indicated statistically significant differences in performance between the Fall 2004 and Spring 2005 assessments. As Table 5 shows, children in the Core Knowledge program made statistically significant gains on every PALS-PreK test and the total score. The significant improvement on PALS is consistent with the observed improvement on the WJ-III. However, PALS is criterion-referenced and more targeted to the curriculum. Therefore, it is more sensitive to changes in specific content knowledge than the WJ-III.

Table 5
Comparison of Pretest and Posttest Performance on the Phonological Awareness Literacy Screening Test – Raw Scores

Achievement Area	<u>Mean Scores</u>		<u>Significance Level</u>
	Fall	Spring	
Rhyme Awareness	3.37	6.52	<.01
ABC Uppercase	7.31	17.55	<.01
Beginning Sounds	2.83	9.53	<.01
Letter Sounds	1.90	8.22	<.01
ABC Lowercase	3.55	12.26	<.01
Print Awareness	3.58	6.69	<.01
Nursery Rhyme	4.68	7.31	<.01
Name Writing	3.63	6.62	<.01
Total PALS	28.35	56.53	<.01

To determine whether the quality of teaching during the preschool year made a difference in terms of how much children gained from Fall to Spring testing, Spring scores were then regressed on Fall scores and Teacher ELLCO scores. As Table 6 shows, children in classrooms with higher quality teaching made greater gains between Fall and Spring than children in classrooms with lower quality teaching.

Table 6
Performance of Students Who Participated in the Core Knowledge Program during 2004-2005 on the Phonological Awareness and Literacy Test in the Spring of the Preschool Year (2005) as a Function Teaching Quality (ELLCO scores), Controlling for PALS Performance in the Fall of 2004

Achievement Area	Quality of Teaching Effect	Significance Level
Rhyme Awareness	Yes	<.01
ABC Uppercase	Yes	<.01
Letter Sounds	Yes	<.01
Beginning Sounds	Yes	<.01
ABC Lowercase	Yes	<.01
Print Awareness	Yes	<.01
Nursery Rhyme	Yes	<.01
Name Writing	Yes	<.01
Total PALS	Yes	<.01

Social Skills Rating Scale

Complete SSRS scores were available on 238 students who attended Core Knowledge in 2004-2005, representing a variety of program sites. For each of the six tests and the two cluster scores, raw scores in the Spring of 2005 were compared to raw scores in the Fall of 2004 using dependent t-tests. For the two component scores (Social Skills and Problem Behavior) raw scores were converted to standard scores; and Fall and Spring performance were also compared using dependent t-tests.. Like IQ the mean score on each SSRS scale at each age is 100.

Table 7
Comparison of Pretest and Posttest Performances of 2004-2005 Core Knowledge Students on the Social Skills Rating Scale Standard Scores

Area Test	<u>Mean Raw Scores</u>			<u>Mean Standard Scores</u>		
	Fall	Spring	Level of Significance	Fall	Spring	Level of Significance
<i>Individual Tests (Raw Scores)</i>						
Cooperation	13.5	14.0	<.05			
Assertion	11.2	12.6	<.01			
Self-Control	12.3	12.9	<.05			
Externalizing	3.4	3.6	NS			
Internalizing	1.0	1.1	NS			
<i>Cluster Scores</i>						
Social Skills	37.3	39.7	<.01	100.4	101.5	NS
Problem Behaviors	4.4	4.7	NS	100.0	101.5	NS

As Table 7 shows, there were statistically significant improvements between Fall and Spring in the scales that assess social skills (cooperation, assertion, and self-control). However, no statistically significant gains were made in the scales tapping problem behaviors (externalizing, and internalizing).

For children in kindergarten and first grade who attended Core Knowledge for the 2002-2003 and 2003-2004 academic years, complete SSRS data were available on 148 for both Fall and Spring. These two groups were analyzed together for purposes of this report. Dependent t-tests were run on the individual tests and cluster scores.

Table 8
Comparison of Pretest and Posttest Performances of combined 2002-2003 and 2003-2004 Core Knowledge Students on the Social Skills Rating Scale Standard Scores during Elementary School

<u>Area Test Scores</u>	<u>Mean Raw Scores</u>			<u>Mean Standard</u>		
	Fall	Spring	Level of Significance	Fall	Spring	Level of Significance
<i>Individual Tests (Raw Scores)</i>						
Cooperation	13.2	14.1	<.01			
Assertion	11.3	12.4	<.01			
Self-Control	13.1	14.1	<.01			
Externalizing	2.1	2.8	<.01			
Internalizing	2.0	2.1	NS			
Hyperactivity	4.1	4.3	NS			
<i>Cluster Scores</i>						
Social Skills	37.6	40.6	<.01	100.0	101.4	NS
Problem Behaviors	8.3	9.2	<.05	99.6	101.1	NS

As Table 8 shows, there were statistically significant improvements between Fall and Spring in the scales that assess social skills (cooperation, assertion, and self-control), just as was observed for children who were currently attending CK. As well there was a statistically significant improvement in externalizing behavior. However, no statistically significant gains were made in the internalizing or hyperactivity scales.

Table 9
Performance of Students Who Participated in the Core Knowledge Program during 2004-2005 on the Social Skills Rating Scale in the Spring of the Preschool Year (2005) as a Function Teaching Quality (ELLCO scores), Controlling for SSRS Performance in the Fall of 2004

Achievement Area	Quality of Teaching Effect	Significance Level
Cooperation	Yes	<.01
Assertion	Yes	<.01
Self-Control	Yes	<.01
Social Skills Cluster Raw Score	Yes	<.01
Social Skills Cluster Standard Score	Yes	<.01
Externalizing	Yes	<.01
Internalizing	Yes	<.01
Problem Behavior Cluster RS	Yes	<.01
Problem Behavior Cluster SS	Yes	<.01

To determine whether the quality of teaching during the preschool year made a difference in terms of how much children improved on the SSRS from Fall to Spring testing, Spring scores were regressed on Fall scores and Teacher ELLCO scores. As Table 9 shows, children in classrooms with higher quality teaching made greater gains between Fall and Spring than children in classrooms with lower quality teaching. Not only were there differences on all individual tests but there were differences in the standard scores for the two cluster scores.

Conclusions

The purpose of this evaluation was to determine whether participation in the Core Knowledge Preschool program results in improved readiness for school and performance after entry into elementary school. In an effort to determine whether the program had a positive impact on children from low-income families, students who participated in the programs were assessed with both standardized norm-referenced and standardized criterion-referenced measures. A series of pre-test – posttest comparisons were made for children who attended CK during the 2004-2005 academic year (Spring 2005 versus Fall 2004) to determine whether gains were made in achievement and social skills. As well pretest – posttest comparisons (Spring 2005 versus Fall 2004) were made students who attended CK during the 2002-2003 and the 2003-2004 academic years.

Granting the limitations on non-experimental designs, results indicate that participation in Core Knowledge had a significant impact on student achievement during the year of the program. There were significant improvements recorded on PALS-PreK scores in all areas of academic functioning and on all the Woodcock-Johnson achievement tests when W scores were analyzed. Importantly, there were significant improvements in standard scores for most areas examined. In terms of the SSRS, positive changes were seen in social skills but not with respect to behavior problems. For all three measures, improvements during the year were linked to the quality of teaching behavior observed (ELLCO scores).

Although it was not feasible to select former CK students (those who attended in 2002-2003 and 2003-2004) in a manner that is fully representative of students for those years, the findings for former students are particularly revealing. Specifically, 2003-2004 students showed W-score gains on every scale from the WJ-III used during their kindergarten year. As well they showed standard score gains on more than half. Likewise, 2002-2003 students showed W-score gains on every test but one during first grade, and showed standard score gains on almost half. This performance from former CK attendees is particularly impressive in view of the declines typically observed in students who come from low-income rural backgrounds. Moreover, CK attendees in kindergarten and first grade also continued to show progress in social skills as measured by the SSRS. There was slightly less evidence of improvement in behavior problems but there were statistically significant gains in externalizing behavior (i.e., the level of externalizing behavior decreased).

Finally, it is important to bear in mind is that there were substantial differences in the fidelity with which CK teachers implemented the program. Some teachers showed a high degree of fidelity, while others did not show such a high level. For every type of outcome used, differences in their quality of teaching was related to children's performance.