The New School Psychology Bulletin
2012, Vol. 9, No.2

Print 2012 ISSN: 1931-793X; Online ISSN: 1931-7948

A Pilot Study of Effects of the Magic Penny Early Literacy Program on Phonemic Awareness and Basic Reading Skills

Carolyn M. Casey
State University of New York at Buffalo

Catherine Cook-Cottone
State University of New York at Buffalo

Courtney N. Baker
The Children's Hospital of Philadelphia

The current pilot study was an investigation into the effectiveness of the Magic Penny Early Literacy Program reading curriculum among kindergarten students. Magic Penny was introduced as a supplement to the existing curriculum in the intervention classroom for approximately three months. Reading achievement was assessed using a selection of tests from the Woodcock-Johnson III Tests of Cognitive Abilities and the Woodcock-Johnson III Tests of Achievement that comprise the Basic Reading Skills Cluster and the Phonemic Awareness III Clinical Cluster. Utilizing a between-group comparison and longitudinal study design, group differences between intervention (n = 19) and comparison (n = 19) classrooms on the early literacy outcome measures were examined using linear regression models. Results provided mixed support for the Magic Penny Early Literacy Program. Analyses indicated that intervention group membership was associated with greater improvement in children's Basic Reading scores, when controlling for pre-test scores. In contrast, intervention group membership was not associated with greater improvement in children's Phonemic Awareness scores. This study represents the first formal evaluation of the Magic Penny Early Literacy Program. However, this pilot study was limited in scope and lacked random assignment. Given its limitations, additional, larger-scale research is warranted to further examine the impact of this new program.

Keywords: Magic Penny, literacy, school-based intervention, phonemic awareness

While the U.S. Department of Education (2002) considers children's literacy the most critical issue in American education, much progress remains to be made in this area. According to the New York State Snapshot Report, 31% of the state's 4th graders fail to reach basic levels of reading achievement (The Nation's Report Card, 2007). Further, despite some recent progress in the scores of ethnic/racial minority students, a reading achievement gap still exists between White and minority children (Lee, Grigg, & Donahue, 2007).

Although early literacy may be one of the most researched topics in education (Solsken, 1993), reading instruction has not evaded controversy. Literacy education has undergone several theoretical shifts during which researchers and educators have argued over the most appropriate method of reading instruction (e.g., whole language versus phonics or direct instruction; Stahl, 2006). In response to the controversies, the United States Congress requested that a panel be created to conduct a comprehensive and rigorous review of research in the area of reading acquisition, with the goal of identifying appropriate, evidence-based practices and disseminating the findings to inform instruction in the schools (National Reading Panel [NRP], 2000; Shanahan, 2006). In an effort to organize the research literature, which included over 100,000 studies, the NRP (2000) decided upon five primary topics around which their report would be centered: phonemic awareness, phonics, vocabulary, fluency, and reading comprehension.

Phonemic awareness, which consists of hearing, identifying, and manipulating phonemes in spoken words (Armbruster, Lehr, & Osborn, 2003; Walpole & McKenna, 2007) falls under a broader language area called phonological awareness (i.e., "alertness to the sounds of spoken language", and one's ability to manipulate those sounds; Joseph, 2006, p. 20). In the area of early literacy, phonological awareness has emerged as a key foundational skill in learning to read (Schuele & Boudreau, 2008), and has even been found to be the strongest predictor of reading performance in some instances (Parrila, Kirby, & McQuarrie, 2004). The National Reading Panel (2000) found that a causal relationship which persists over time exists between phonemic awareness training and student improvement in the areas of reading and spelling (NRP, 2000). Among the most successful characteristics of phonemic awareness training were explicit instruction in phoneme manipulation using letters (focusing on only one or two manipulations at a time), and small-group instruction (NRP, 2000). It is important to note that, despite the positive results in this area, the panel cautioned that phonemic awareness training alone is insufficient as a reading instructional strategy; additional techniques must be included for a reading program to be considered comprehensive (NRP, 2000).

Unlike phonological awareness and phonemic awareness, which focus on alertness to sounds within words and phonemes within words respectively, phonics involves learning specific letter-sound relationships that facilitate children's ability to read and spell (Gillon, 2004). The instructional methods used to teach phonics have traditionally been considered fairly tedious, based on the notion that many programs relied heavily on rules and structured worksheets during training (Stahl, Duffy-Hester, & Dougherty Stahl, 2006). However, given the important role it plays in reading (Dombey, 1999), phonics has remained a focus of research. In the area of phonics, the National Reading Panel (2000) investigation found phonics instruction to be an effective practice for students in kindergarten through sixth grade, including children with reading difficulties. When examining the potential differential impact of phonics instruction by grade level, the report found that those in kindergarten and first grade were most positively impacted; this finding led the panel to recommend instruction targeted toward these age groups specifically (NRP,

2000). Among the different approaches to phonics instruction, the NRP (2000) identified synthetic phonics as effective for children with reading difficulties and for children from lower socioeconomic backgrounds. Synthetic phonics is an instructional approach that begins with letter identification, and is followed by learning each letter's corresponding sound, and how to combine those sounds into words (NRP, 2000). Based on their review, the NRP (2000) emphasized the important role phonemic awareness plays in phonics instruction. Children may be able to learn basic letter-sound relationships (i.e., phonics), but they must also learn how to apply that knowledge in order to become successful readers. Therefore, the NRP (2000) recommended coupling phonics instruction with phonemic awareness tasks like blending (i.e., combining phonemes to make words; Armbruster et al., 2003) and segmentation (i.e., separating words into phonemes; Armbruster et al., 2003).

Although the current study will focus on the development of key early literacy skills related to phonemic awareness and phonics, it is important to note that the NRP (2000) also reviewed important findings regarding vocabulary development, reading fluency, and reading comprehension. Broadly, age-appropriate vocabulary instruction (e.g., pre-learning of words, task restructuring, and repeated exposure) was correlated with increased levels of reading comprehension (NRP, 2000). In the fluency domain, the technique of repeated guided oral reading was associated with improvement in the areas of word recognition, fluency, and reading comprehension across education settings (i.e., general and special education), grade levels, and students with varying skill levels. As a secondary fluency technique, the use of independent silent reading failed to demonstrate a positive impact on reading achievement and fluency, mainly because most studies in this area were correlational rather than experimental. Finally, the NRP (2000) reviewed research in the area of reading comprehension, and outlined seven reading comprehension strategies that showed promising results in their review: monitoring (i.e., students' self-awareness of what they do and do not understand in the text; Armbruster et al., 2003); cooperative learning (i.e., when students work together to identify and practice strategies for comprehending text; NRP, 2000); using graphic or semantic organizers (i.e., visual displays that depict key concepts within a reading; Armbruster et al., 2003); answering questions (i.e., responding to key questions that can help structure the reading); generating questions (i.e., creating questions about the story to which students can also respond); story structure (i.e., having students rely on the story's structure to help them answer comprehension questions; NRP, 2000); and summarization (i.e., having students synthesize the reading into main ideas). Individually, each of these strategies has shown some degree of success in the classroom (NRP, 2000). However, the NRP (2000) findings indicated that teaching *multiple* text comprehension strategies is most effective.

Significance of the Present Study

Although the National Reading Panel (2000) has outlined several key components of a solid reading program, according to Casey and Howe (2002), many current basal reading programs fail to utilize research-based principles and skills that are necessary for literacy acquisition. As a result of this gap between research and practice, school psychologists are urged to advocate for the use of empirically-based reading programs which have proven to be successful. One of the problems, however, is the dearth of empirically supported reading programs. The present study aims to fill this deficit by investigating the effectiveness of an early literacy intervention for kindergarten students that was specifically designed based upon empirical findings from the National Reading Panel (NRP; 2000), along with additional research in the area of reading acquisition.

Current Intervention: The Magic Penny Early Literacy Program

The Magic Penny Early Literacy Program was created based upon the findings of the National Reading Panel (2000), along with additional research in the area of reading acquisition. Magic Penny is a literacy program that uniquely combines evidence-based techniques, primarily focusing on the development of phonemic awareness skills. At the early literacy level, the goals of Magic Penny are to teach children how to process language through hearing and help them understand that sounds are represented by symbols (S. Schneider, personal communication, October 4, 2009). This program can be implemented in conjunction with any basal reading curriculum (e.g., Macmillan/McGraw-Hill, Scott Foresman, Houghton Mifflin, or Harcourt; Schneider &Schneider, n.d.), and typically consists of 20–30 minutes of daily instruction. Instruction begins with a foundation in phonemic awareness, and moves sequentially through reading and writing, then understanding what is read and developing critical thinking skills (Sciarrino, 2009). In the first level of the Magic Penny Early Literacy sequence (Level A: Phonemic Awareness Skills), children learn phonemic awareness skills by identifying different sounds in words (e.g., beginning, middle, and ending sounds), rhyming, manipulating sounds (e.g., removing and adding sounds from words to make new words), blending and segmenting sounds, and identifying words (Sciarrino, 2009). Once phonemic awareness is mastered, children begin activities focused on letter-sound relationships, how to form letters that represent sounds from which words can be created, how to decode or blend sounds into words, and how to segment words into sounds (Sciarrino, 2009). It is important to note that children are not taught to memorize sight words (defined as words that are recognized as a whole; Joseph, 2006), or guess the correct word while reading based on context; instead, children learn to decode *each* sound within a word (Schneider & Schneider, 2009b). All of these early literacy skills are taught through the use of interactive games and developmentally appropriate activities, like matching and rhyming activities using brightly colored picture cards.

A unique feature of Magic Penny is its incorporation of parents into literacy development by providing materials that can be used at home. This is consistent with findings from Senechal's (2006) meta-analysis of family literacy interventions, which found parental involvement to be positively associated with children's reading development (d = 0.68). Consistent with this recommendation, Magic Penny supports parents who are willing to participate in program activities at home. For example, at the beginning level, the program includes Parent Kits with an instructional DVD about Magic Penny and explicit instructions for completing the

literacy activities at home. This aligns with Senechal's (2006) recommendation that educators provide parents with appropriate training techniques. In addition to providing home materials, the developers of Magic Penny also offer parent workshops over the course of the school year to provide additional support.

Early Findings

Preliminary data from participating school districts suggests that Magic Penny has positive results. The first elementary school to implement Magic Penny saw a 48% increase in grade four New York State English Language Achievement (ELA) proficiency results when compared to the previous class of students who did not receive Magic Penny instruction (Schneider & Schneider, 2009a). At that same school, 100% of children who received the Magic Penny Early Literacy Program in kindergarten met or exceeded New York State proficiency standards on the third grade ELA exam in the 2005-2006 school year (Schneider & Schneider, 2009a).

In 2008, the Magic Penny curriculum was expanded in another New York school to target universal pre-kindergarten children. Based on the Magic Penny Phonological Assessment tool, which was developed by the Magic Penny Early Literacy Institute to evaluate student progress, scores across the three time points during the year—September, January, and June—indicated that the June, 2009 average pre-kindergarten score was 117% higher than the average score for kindergarten students who entered school in September, 2008 without receiving Magic Penny during pre-kindergarten (Schneider & Schneider, 2009a).

During the 2008–2009 academic year, another participating school was awarded the Exemplary Reading Program Award by the International Reading Association for its Magic Penny implementation (Schneider & Schneider, 2009a). Most recently, an additional New York elementary school completed its first year of Magic Penny implementation for all kindergarten and first grade students, and compared the school's performance on curriculum-based measures, including AIMSweb early literacy probes, to the performance of other schools in the district. After one year of implementation, the school using Magic Penny improved performance from below district standards to above district standards on Phoneme Segmentation Fluency (S. Wolf, personal communication, June 16, 2009).

Despite the accolades this program has received from the schools, a systematic study had yet to be conducted on the Magic Penny program. Therefore, the aim of the current study was to conduct the first official investigation into the effects of Magic Penny on key early literacy indicators. It was hypothesized that the program would increase Phonemic Awareness and Basic Reading skills, measured as proxies of two major foci—phonemic awareness and phonics—of the NRP (2000) report.

Hypotheses

In the current pilot study, hypothesis one was that intervention group membership would be associated with greater improvement in children's Basic Reading scores. Hypothesis two was that intervention group membership would be associated with greater improvement in children's Phonemic Awareness scores.

Method

Participants

Participants from this study were drawn from a public elementary school in the state of New York in the United States that was chosen for the study for the following reasons: (a) the students had no prior exposure to the Magic Penny Program, and (b) the teachers had been introduced to Magic Penny over their summer school sessions and expressed interest in the program. Therefore, the school was asked to participate in a research study on program effectiveness. This public elementary school includes approximately 360 students in grades kindergarten through five. The majority (81%) of students in this population were Caucasian. Forty-one percent of the student body were reported to be eligible for free or reduced lunches (New York State Report Cards, n.d.). Study participants included students from two kindergarten classrooms (n = 38; ages 4–6 years old) who were enrolled for the 2009–2010 academic year. Prior to the beginning of the academic year, one classroom teacher volunteered and was trained by the director of the Magic Penny Early Literacy Institute to implement this program as a supplemental reading intervention; therefore, the reading achievement of the students in the intervention classroom (n = 19) was compared to the reading achievement of the students in a separate kindergarten classroom (n = 19), which continued to implement the school's existing reading curriculum, Houghton Mifflin Reading (Houghton Mifflin, 2002) for the duration of the study period. This group of children who did not receive Magic Penny instruction is referred to as the "comparison group," which, in this study, denotes that these children received "treatment as usual" in this educational setting. Magic Penny training was delivered exclusively to the intervention classroom teacher; the comparison classroom teacher was not provided with any instruction or materials related to this program.

Procedures

After obtaining parental consent, all children were given five phonemic awareness/reading tests from the standardized Woodcock-Johnson III Tests of Cognitive Ability and the Woodcock-Johnson III Tests of Achievement. This pre-testing occurred during the first ten days of the academic year. Overall, the assessment took approximately 30 minutes per student, and all student responses were recorded directly on the Woodcock-Johnson III standardized test protocols. All protocols were scored using the Woodcock-Johnson III Compuscore and Profiles Program, which is a computer program that automatically scores the tests and produces score reports.

For the duration of the study period, the intervention classroom teacher implemented the Magic Penny Early Literacy Program for approximately 20–30 minutes daily for all students. This is consistent with the National Early Literacy Panel (2008) recommendation that reading evaluations assess the effects of early literacy programs implemented by classroom teachers, not researchers. The founder of the Magic Penny Early Literacy Institute was available to provide ongoing professional development or support for the intervention teacher when needed. Further, two parent workshops were provided by the founder of the Magic Penny Early Literacy Institute. One was conducted at

the beginning of the school year in September to provide parents with Magic Penny parent materials and teach them how to play Magic Penny early literacy games at home with their children. The second workshop was conducted in November to provide parents with additional training and support as they worked with their children to foster literacy development at home. During this second workshop, parents were also invited to visit the intervention classroom to observe their children participating in Magic Penny games and activities.

In December, prior to the winter recess, all kindergarten children again completed the Woodcock-Johnson III tests as a post-test measure. After the post-test data was collected in December, the other kindergarten classroom teacher (whose children acted as comparison subjects for the purposes of this study) was given the opportunity to start incorporating the Magic Penny Early Literacy Program for the remainder of the year. In this way, if Magic Penny was deemed successful in the intervention classroom over the first half of the school year, the remaining kindergarten children would be able to benefit from the program for the second half of the year. Although this option was available to the comparison classroom teacher, this teacher did not choose to begin incorporating Magic Penny into the comparison classroom curriculum.

Instruments

All data were collected by supervised graduate-level school psychology students who had undergone extensive training in cognitive and achievement assessment through their academic program of study. The Woodcock-Johnson III assessment battery was selected based on its strong psychometric properties, and the fact that it provides reliable selective testing methods (Schrank, Flanagan, Woodcock, & Mascolo, 2002), which makes it possible to assess phonemic awareness and basic reading skills specifically.

Woodcock-Johnson III Tests of Cognitive Abilities (WJ III Cognitive; Woodcock, McGrew, & Mather, 2001b). The WJ III Cognitive test was created based on the Cattell-Horn-Carroll (CHC) theory of intelligence, and consists of 20 total tests (10 from the standard battery, and 10 from the extended battery) that assess a range of cognitive abilities and form clusters that help provide important interpretive information for educational or diagnostic purposes (Mather & Woodcock, 2001b). Among these clusters is General Intellectual Ability, along with the Verbal Ability, Thinking Ability, and Cognitive Efficiency performance clusters. Further, the WJ III Cognitive also provides cluster scores for a range of additional cognitive categories, CHC factors, and clinical clusters (Mather & Woodcock, 2001b). According to Mather and Woodcock (2001b), the WJ III Cognitive lends itself to use as an outcome measure in research or program evaluation.

All cluster and test scores have a mean of 100 and a standard deviation of 15 (Mather & Woodcock, 2001b). For the purposes of the current research study, all children were given two Phonemic Awareness tests from this battery—Sound Blending and Incomplete Words—that assess auditory processing. Sound Blending requires children to listen to a combination of phonemes and blend those phonemes into a recognizable word (Mather & Woodcock, 2001b). Its median split-half reliability for children ages 5-19 is .86. Comparatively, Incomplete Words requires children to listen to words with missing phonemes, and recognize the words. Its median split-half reliability among children ages 5-19 is .77 (Mather & Woodcock, 2001b).

Woodcock-Johnson III Tests of Achievement (WJ III Achievement; Woodcock, McGrew, & Mather, 2001a). The WJ III Achievement test was co-normed with the WJ III Cognitive and consists of 22 total tests (12 from the standard battery and 10 from the extended battery) that assess a range of academic skills (e.g., reading, mathematics, written language, oral language, and academic knowledge; Mather & Woodcock, 2001a). As with the WJ III Cognitive, the WJ III Achievement can be used for educational, diagnostic, and research purposes. Together, the WJ III Cognitive and Achievement were co-normed on over 8,000 children and adults ages 2 through more than 80 years of age, and represent comprehensive measures "designed to provide the most valid methods for determining patterns of strengths and weaknesses based on actual discrepancy norms" (McGrew & Woodcock, 2001, p. 4).

Consistent with the WJ III Cognitive, all WJ III Achievement test and cluster scores have a mean of 100 and a standard deviation of 15. Because the current study has a primary focus on reading achievement, students were given the following three relevant tests from this battery: Letter-Word Identification, Word Attack, and Sound Awareness. Letter-Word Identification assesses children's abilities to first identify letters in isolation, and then identify a series of words. Its median split-half reliability among children ages 5–19 is .91 (Mather & Woodcock, 2001a). In contrast to Letter-Word Identification, Word Attack requires children to first identify letter sounds in isolation, and then pronounce a series of low-frequency and nonsense words. Its median split-half reliability for children ages 5–19 is .87. Finally, Sound Awareness assesses a range of phonological awareness skills: rhyming (i.e., children's ability to identify and then produce words that rhyme with a target word), deletion (i.e., children's ability to remove parts of, or sounds from, words to make new words), substitution (i.e., requiring children to substitute words, word endings, or letter sounds to create new words), and reversal (i.e., requiring children to reverse the order of compound words, and then letter sounds within words, to create new words; Mather & Woodcock, 2001a). The median split-half reliability of the Sound Awareness test is .81 for children ages 5–19.

Together, the selection of WJ III Cognitive and Achievement tests chosen for the present study comprised two overall cluster scores: the Phonemic Awareness III Clinical Cluster comprised of the Sound Blending, Incomplete Words, and Sound Awareness tests (cluster median split-half reliability = .91; Schrank et al., 2002) and the Basic Reading Skills Cluster comprised of the Letter-Word Identification and Word Attack tests (cluster median split-half reliability = .95; Mather, Wendling, & Woodcock, 2001). These two clusters were chosen because they assess the phonemic awareness and phonics skills that the NRP (2000) has identified as critical to reading acquisition.

Analytic Approach

The purpose of this study was to examine the effectiveness of a new early literacy program, utilizing between-group comparison and a longitudinal study design. More specifically, the current study examined group differences between intervention and comparison groups on early literacy outcome measures (i.e., Basic

Reading Skills and Phonemic Awareness, as measured by the WJ III), controlling for participants' pre-test performance. Linear regression models were fit to test the study hypotheses. In the first regression model, group differences between Basic Reading Skills Cluster post-test scores were assessed, controlling for children's Basic Reading Skills Cluster pre-test scores. In a second regression model, group differences between Phonemic Awareness III Clinical Cluster post-test scores were assessed, controlling for children's Phonemic Awareness III Clinical Cluster pre-test scores. In an effort to better understand these findings, additional exploratory analyses were also conducted, adding age and gender as additional predictor variables to the regression models.

Results

Descriptive Statistics

The original study sample size included 41 kindergarten children (20 from the intervention classroom, receiving Magic Penny instruction and 21 from the comparison classroom, receiving Houghton Mifflin instruction alone). During pre-testing, two children (one from the intervention classroom and one from the comparison classroom) chose to discontinue their participation and did not complete pre-testing; therefore, they were excluded from the study. Further, one additional child was identified as having a severe speech impediment that interfered with the examiner's ability to comprehend his answers; he did not complete pre-testing, and was also excluded from the study. Given this, the final sample size was 38 (19 children from the intervention classroom, and 19 children from the comparison classroom). Of these 38 participants, 22 were female, and 16 were male. The average age at pre-test was 63.76 months (SD = 3.80). Given that random assignment was not possible, pre-test differences on outcome variables were assessed. No significant differences were found between the intervention and comparison groups at pre-test on Basic Reading Skills (intervention: M = 105.37, SD = 13.17, comparison: M = 110.21, SD = 12.75) or Phonemic Awareness (intervention: M = 98.53, SD = 11.25, comparison: M = 102.00, SD = 12.69).

Descriptive statistics for and intercorrelations between study variables are presented in Table 1. Of note, the observed mean scores on the outcome variables in this sample fell within the "average" WJ III classification range (i.e., scores between 90 and 110, based on the WJ-III standard score norms with a mean of 100 and a standard deviation of 15; Mather & Woodcock, 2001a). The one exception to this is that the mean Basic Reading Cluster post-

Table 1.

Descriptive Statistics for and Intercorrelations Between the Predictor Variables

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1. Group ($N_{Comparison}$ = 19 (1); $N_{Intervention}$ = 19 (2))	-----						
2. Gender (N_{Male} = 16 (1); N_{Female} = 22 (2))	.11	-----					
3. Age (M = 63.76 mo, SD = 3.80 mo)	-.11	-.04	-----				
4. Basic Reading Pre-test (M = 107.79, SD = 13.02)	-.19	.08	.14	-----			
5. Basic Reading Post-test (M = 122.61, SD = 10.33)	.06	.14	-.34*	.80***	-----		
6. Phonemic Awareness III Pre-test (M = 100.26, SD = 11.96)	-.15	.19	-.22	.50**	.57***	-----	
7. Phonemic Awareness III Post-test (M = 108.66, SD = 12.86)	-.10	.29+	-.32*	.50**	.59***	.87***	-----

Note. + p < .10, * p < .05, ** p < .01, *** p <. 001

test score for the sample (M = 122.6) fell within the "superior" WJ III classification range (i.e., scores between 121 and 130). As expected, an examination of intercorrelations between the outcome variables showed that pre-test outcomes were significantly correlated with post-test outcomes; significant correlations also existed between the two measures—Basic Reading Skills and Phonemic Awareness. All statistically significant correlations are of medium and large effect sizes. For means and standard deviations for scores on Basic Reading and Phonemic Awareness III as a function of group, see Table 2. As reflected in Table 2, both the intervention and the comparison groups' scores improved from pre-test to post-test on each outcome measure.

Primary Outcomes

Basic Reading Skills Cluster. Consistent with the first hypothesis, intervention group membership was associated with greater improvement in children's Basic Reading scores (see Figures 1 and 2). On average, children in the intervention classroom improved 4.42 standard score points more from pre-test to post-test than children in the comparison classroom, controlling for pre-test scores, b = 4.42, 95% CI = [.39, 8.46], se = 1.99, p = .03; intervention: M = 123.21, SD = 10.60; comparison: M = 122.00, SD = 10.30. This model accounted for 67.86% of the variance in post-test scores, $F(2, 35) = 36.95$, $p < .001$. The effect size associated with this model is considered large, f^2 = 2.12. Further, group membership explained significantly more of the variance when added; in fact, it explained 4.5% of the variance in Basic Reading performance above and beyond the effects of children's Basic Reading pre-test scores, $F(1, 35) = 4.95$, p = .03. This is considered a medium sized effect, f^2 = .14.

Phonemic Awareness III Clinical Cluster. Contrary to the second hypothesis, intervention group membership was not associated with greater improvement in children's Phonemic Awareness scores, b = .81, se = 2.10, p = .70; intervention: M = 107.42, SD = 15.10; comparison: M = 109.89, SD = 10.41 (see Figures 3 and 4).

Additional Analyses

In an effort to better understand these findings, post-hoc exploratory analyses were conducted utilizing age and gender as additional predictor variables in the regression models. No significant group differences between boys and girls were found on Basic Reading Cluster scores, b = 1.09, se = 2.02, p = .59, or Phonemic Awareness III Clinical Cluster scores, b = 3.28, se = 2.12, p = .13. Conceivably, the intervention may have worked differentially for boys and girls (e.g., perhaps one gender may have responded more favorably to the interactive nature of the curriculum). However, controlling for pre-test scores, the relationship between post-test scores and group membership did not differ between boys and girls on Basic Reading, b = -.22, se = 4.16, p = .96, or Phonemic Awareness, b = 6.02, se = 4.12, p = .15.

The authors also tested whether age was a significant predictor of either Basic Reading Cluster post-test or Phonemic Awareness III Clinical Cluster post-test scores. Therefore, additional regression models were fit including age as a predictor variable. Controlling for pre-test scores and intervention group, age was significantly negatively associated with Basic Reading post-test scores, such that older children scored lower than younger children on average, b = -.56, 95% CI = [-1.07, -.04], se = .25, p = .03. The regression weight associated with group membership was reduced slightly in magnitude once age was added to the model, b = 3.85, 95% CI = [-.02, 7.72], se = 1.90, p = .05. The effect size associated with this model is considered large, f^2 = 2.56. Comparatively, there was no significant relationship between the Phonemic Awareness III Clinical Cluster post-test scores and age, b = -.45, se = .28, p = .12. It was also assessed whether age moderated the relationship between intervention group membership and post-test scores. This was not supported; the slope of the relationship between age and post-test scores did not differ by group on Basic Reading, b = -.33, se = .50, p = .52 or Phonemic Awareness, b = -.91, se = .55, p = .11, when controlling for pre-test scores.

Discussion

Although the Magic Penny Early Literacy Program is currently implemented in many schools across the country, it has yet to undergo a systematic evaluation of its effectiveness. Therefore, the current study was the first to assess the effectiveness of this new program, using standardized reading assessment measures among a sample of kindergarten students. The primary results of this evaluation provided partial support for the Magic Penny Early Literacy Program's effectiveness. Over the course of the study period, children in the intervention classroom showed significantly more improvement on a measure of Basic Reading Skills (which assesses children's sight-word recognition and phonics knowledge; Mather et al., 2001), when compared to children in the comparison classroom. In fact, the intervention classroom improved almost one third of a standard deviation more than the comparison group on this measure.

Unexpectedly, no significant performance differences were detected on the Phonemic Awareness III Clinical Cluster outcome measure, which assesses auditory processing and phonological awareness (Schrank et al., 2002). This finding is particularly surprising, given that the Magic Penny Early Literacy Program is based upon teaching a strong foundation of phonemic awareness skills. In the past, Magic Penny developers have identified performance differences in phonemic awareness among children who received Magic Penny instruction compared to those who did not receive such instruction (Schneider & Schneider, 2009a). However, these findings were based upon the developers' own outcome measure, the Magic Penny Phonological Awareness Assessment (Schneider & Schneider, 2008). To the authors' knowledge, this Magic Penny assessment has not undergone any validation studies; therefore, it has no documented reliability or validity indices. Given the close alignment between the Magic Penny assessment and the Magic Penny intervention, the program's own measure may have been more likely to detect effects of the Magic Penny program. If so, it is possible that the developers' Magic Penny Phonological Awareness Assessment over-estimated changes in phonemic awareness as a result of the Magic Penny intervention.

It is also important to note that there appear to be differences in content between the Magic Penny Phonological Awareness Assessment and the tests that comprise the WJ III Phonemic Awareness Clinical Cluster. More specifically, the Magic Penny

Table 2.

Means and Standard Deviations of the Outcome Variables at Both Time Points, by Group

	Basic Reading Cluster		Phonemic Awareness III Clinical Cluster	
Group	Pre-test M (SD)	Post-test M (SD)	Pre-test M (SD)	Post-test M (SD)
Intervention	105.37 (13.17)	123.21(10.60)	98.53 (11.25)	107.42 (15.10)
Comparison	110.21 (12.75)	122.00 (10.30)	102.00 (12.69)	109.89 (10.41)

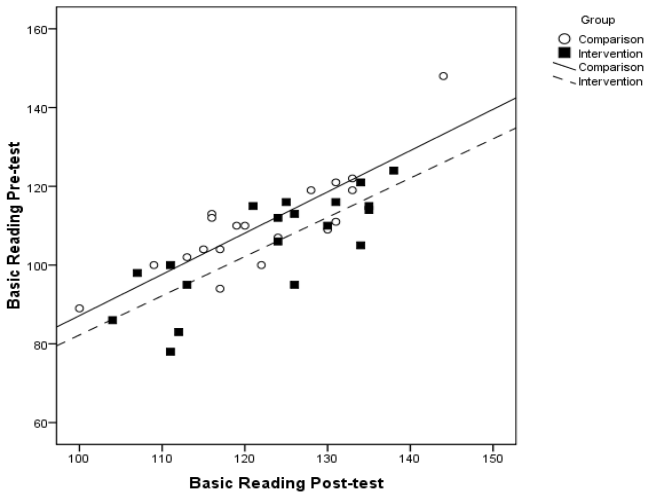
measure includes targets such as the assessment of concepts of spoken word (i.e., whether students understand that sentences are composed of individual words), along with beginning and ending sound recognition (Schneider & Schneider, 2008). Further, the Magic Penny Phonological Awareness measure tests for phoneme segmentation and decoding ability (i.e., whether children can read a list of basic words). None of these skills is directly assessed in the WJ III Phonemic Awareness measure. In fact, decoding ability is included as a skill that falls under the WJ III Basic Reading Skills Cluster. Given these observations, it appears that the Magic Penny Phonological Awareness measure has unique characteristics that do not align with the WJ III Phonemic Awareness Clinical Cluster. These content differences may help explain the discrepancy in findings between past Magic Penny program effects, and those found in the current evaluation. In an effort to further explain this unexpected non-significant finding on the Phonemic Awareness III Clinical Cluster, analyses were run using individual test total raw scores to examine whether there was a different pattern of results across the tests that comprise this cluster score. However, there were no significant group differences on any individual test.

To better understand the process of change associated with the Magic Penny intervention, the roles of children's gender and age were investigated. Specifically, it was first assessed whether these characteristics were associated with children's outcomes, and then whether these characteristics were differentially associated with change in outcome scores as a result of participating in the intervention. Contrary to hypotheses formed during exploratory analyses, gender was not demonstrated to be associated with child outcomes in this intervention. In contrast, results from regression analyses indicated that age was significantly negatively associated with Basic Reading Skills post-test scores, such that older children scored lower than younger children. As further evidence, a correlational analysis revealed a negative and significant correlation between age and Basic Reading post-test scores. This was an unexpected finding, especially given the narrow age range of the sample (from 57 months to 72 months). Because the WJIII includes age as a factor in the standardization of the test, it was hypothesized that the test norms may have been playing a role in this outcome (i.e., the norms may have become increasingly stringent over the period of months in this age range). Therefore, raw scores for the two tests that comprise the Basic Reading Skills Cluster (i.e., Letter-Word Identification and Word Attack) were examined to test the possibility that the standardization procedures were impacting this pattern of results. When raw scores were analyzed, there was no significant correlation between age and Letter-Word Identification post-test scores. Further, a positive and significant correlation emerged between age and Word Attack post-test scores, such that older children scored higher than younger children, which is a result that is more in line with expectations. Overall, this suggests that the initial significant negative association between age and Basic Reading Skills post-test scores may have been a reflection of the WJ III standardization procedures. Future research that includes a more diverse age range may help to further clarify how age may be related to Magic Penny outcomes, or whether the current findings related to age are simply artifacts of this study sample.

Limitations

A major strength of this study was the fact that it was conducted in real classroom settings. However, because this study did take place in a natural school environment, there were several inherent limitations that pose threats to this study's internal and external validity. Internal validity addresses the degree to which the study outcomes can be attributed to the intervention in question (Trochim & Donnelly, 2007). In this real world school-based study, there are several confounding factors that may have played a role in the study results. First, it was noted that the intervention classroom had access to a part-time teacher's aide who assisted with learning activities for a portion of the school day. In contrast, the comparison classroom had access to a full-time teacher's aide. It is possible that this difference could have confounded the current findings (e.g., the comparison classroom children may

Figure 1. Basic Reading Cluster Scores by Group.



have received more, or meaningfully different, support with the addition of a second adult in the room throughout the school day).

In the current study, no intervention fidelity measures were used to assess the intervention classroom teacher's implementation of the Magic Penny protocol. Although the teacher was well-trained and had access to consultation with the developer of the program, because intervention fidelity was not monitored, it is possible that the classroom teacher did not follow all of the Magic Penny protocols as directed. If the program was not implemented as intended, it is possible that a lack of significance was not due to the ineffectiveness of the intervention in general, but rather to the way that the intervention was carried out in the context of this study; similarly, any significant results may have been attributed to instructional methods unrelated to the Magic Penny curriculum (e.g., individual teaching style).

One of the innovative aspects of the Magic Penny Early Literacy Program is the home-based parent involvement component. However, the scope of this pilot study restricted researchers' ability to examine the degree of parental involvement, or even whether parents were actually implementing Magic Penny activities in the home, and as directed. Accessing this information (e.g., by asking parents to report their levels of involvement at home) could help researchers identify whether the Magic Penny Early Literacy Program's unique addition of parent materials adds an important ingredient to literacy development, and how parents' intervention fidelity may impact their children's reading achievement.

In contrast to internal validity, external validity refers to how well the results from a research study can be generalized to broader populations in different contexts (Trochim & Donnelly, 2007). One of the primary threats to external validity is a direct result of the fact that it is an effectiveness trial. That is, in this study, random assignment to groups was not possible. Although analyses confirmed that the two groups' pre-test scores were not statistically different at the start of the study period, at least with regard to one outcome measure, the trend was for the comparison group to have better pre-test scores than the intervention group. The directionality of this trend works against the study hypotheses, and it is possible that it made it harder to see change in children's scores between the two groups. Without random assignment, it is more difficult to infer a causal relationship between the intervention and the study outcomes. This problem was compounded in the context of this pilot study, because of the small sample size. Not only is sample size a key component in the computation of statistical power, or the chance that an intervention effect will be detected if it is truly there (Trochim, 2006), but this pilot study was limited to only two classrooms. Without utilizing multiple classrooms in each condition, effects related to the effectiveness of each teacher or the composition of each classroom could be playing an important role that this study is unfortunately unable to assess.

It is also important to note that a regression model can only explain outcomes based on the factors included in the model (i.e., more informally, it can only tell us what we put into it). Other potentially important factors (e.g., teachers' level of experience, student socioeconomic status, and kindergarten readiness levels) may have contributed to the pattern of results. However, the current study's author did not have access to this information. Because only data on age and gender were collected, the potential impact of any additional predictor variables is unknown.

A final limitation of the current study is the duration of the study period, which was only three months (between September and December of the participants' kindergarten year). This short study period was originally intentional; during the recruitment process, there were some concerns about withholding the reading intervention from other kindergarten classroom teachers who may have been interested in utilizing its program components in their instruction. Therefore, in order to elicit participation from a comparison classroom, the consent forms stated that, if the intervention program were found to be effective over the course of the first three months of the school year, the comparison classroom would have the opportunity to incorporate Magic Penny into its curriculum for the second half of the school year. Making compromises like this is a consequence of effectiveness work,

Figure 2. Alternative view of Basic Reading Cluster scores by group.

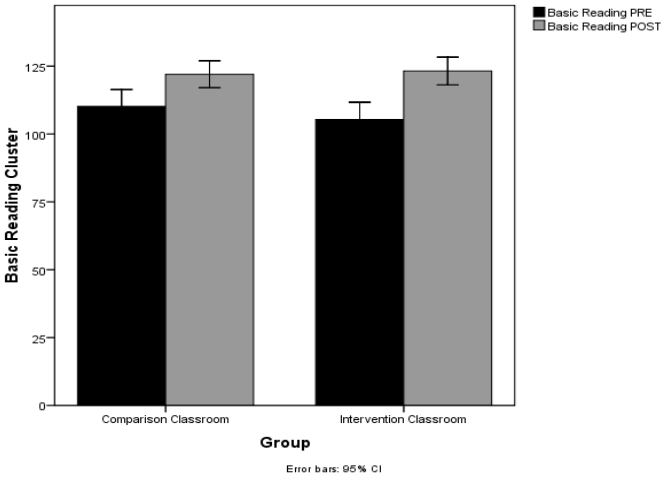
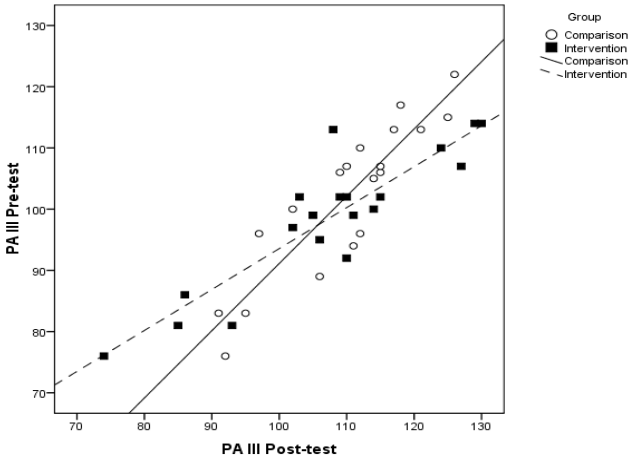


Figure 3. Phonemic Awareness III Cluster Scores by Group.



where the needs of the school must be balanced with the needs of the research. Although the study period was brief, the findings associated with the Basic Reading Skills Cluster suggested that intervention children improved more than comparison children over time, when pre-test performance was controlled. Despite this, it is difficult to know if the improvement would be large enough to translate into a clinically significant impact on performance in the classroom. However, this finding may be tied into the fact that the study period was so brief. It is possible that even stronger effects on Basic Reading, and significant effects on Phonemic Awareness may have emerged if the children had been exposed to the program for a longer duration.

Implications for Future Research

Despite its limitations, the current pilot study is the first to formally evaluate the effectiveness of the Magic Penny Early Literacy Program on reading outcomes, using standardized reading assessment measures. Results from the current evaluation suggest that further research is needed to more clearly understand the efficacy of this program. More specifically, future effectiveness studies would benefit from addressing many of the areas of limitations previously discussed. Further, because the current study only utilized outcome measures related to two of the five areas (i.e., phonemic awareness and phonics) outlined in the NRP (2000) report, additional studies should consider incorporating measures that examine how Magic Penny influences vocabulary, fluency, and comprehension development.

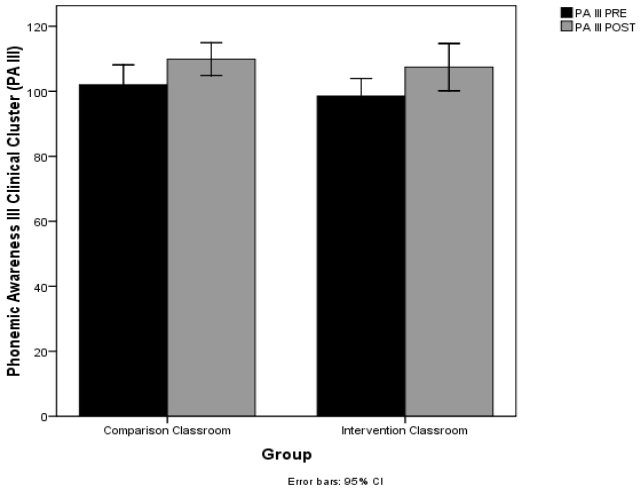
Despite the challenges inherent in doing research in school settings, use of random assignment in future evaluations should be utilized at the child, classroom, and/or school level. Given the substantial resources dedicated to a research study, randomly assigning interventions provides considerable benefit or value, helping rule out the many confounds often used to discredit school-based research. Future effectiveness studies may also benefit from examining the role of parental involvement in the Magic Penny Early Literacy Program, and how this may affect literacy development. It is also advised that additional research consider utilizing a more extensive longitudinal design to examine the impact of the Magic Penny curriculum on reading achievement across the elementary school grades (e.g., to assess whether positive outcomes persist across time, or whether intervention children reach a ceiling and fail to show long-term treatment effects as they move through subsequent grades).

In addition to addressing some of the limitations of the current study, it is recommended that future research incorporate curriculum-based measures (CBMs) such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) or AIMSweb probes as additional indicators of early literacy skill acquisition. A strength of the current study is its use of standardized and highly reliable outcome measures. However, there has been a recent shift in education toward the use of more informal school-based assessments, or curriculum-based measures, that provide a direct link between assessment and instruction (Howell, Kurns, & Antil, 2002). Further, these assessments may be more feasible for use in reading evaluation studies; they take little time to administer, and they are typically given to all students at the elementary school level. Their ease of administration, coupled with the fact that all students should have CBM scores recorded at regular time points throughout the year, could potentially facilitate both child improvement in language and literacy and school psychology research on these processes.

Conclusion

The current study is a unique addition to the literature on evidence-based reading interventions, as it represents the first formal evaluation of the Magic Penny Early Literacy Program in the context of a real classroom setting. This program has demonstrated initial success in the area of basic reading skills among a sample of kindergarten students, but has failed to show clear evidence of improvement in the area of phonemic awareness. This pilot study was limited by a small sample size and lack of random assignment. Additional larger-scale research is warranted to further examine the impact of this potentially

Figure 4. Alternative view of Phonemic Awareness III Cluster scores by group.



promising new program. Specifically, future studies would benefit from incorporating a blend of standardized and curriculum-based early literacy measures, and would be further strengthened by monitoring intervention fidelity and examining the Magic Penny curriculum's impact on reading achievement over time through the use of a more extensive longitudinal design.

References

Armbruster, B. B., Lehr, F., & Osborn, J. (2003). *Put reading first: The research building blocks For teaching children to read: Kindergarten through grade 3*(2nd Ed.). National Institute for Literacy. Retrieved from http://nifl.gov/partnershipforreading/publications/PFRbookletBW.pdf

Casey, A., & Howe, K. (2002). Best practices in early literacy skills. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 721-735). Washington, DC: National Association of School Psychologists.

Dombey, H. (1999). Towards a balanced approach to phonics teaching. *Reading, 33*(2), 52-58.

Gillon, G. T. (2004). *Phonological awareness: From research to practice*. New York, NY: Guilford Press.

Houghton Mifflin. (2002). Introduction. Retrieved from http://www.eduplace.com/marketing/nc/pdf/fw_p02-03.pdf

Howell, K. W., Kurns, S., & Antil, L. (2002). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 753-771). Washington, DC: National Association of School Psychologists.

Joseph, L. M. (2006). *Understanding, assessing, and intervening on reading problems: A guide for school psychologists and other educational consultants*. Bethesda, MD: National Association of School Psychologists.

Lee, J., Grigg, W. S., & Donahue, P. L. (2007). The nation's report card: Reading 2007 executive summary. National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2007/2007496.asp#pdflist

Mather, N., Wendling, B. J., & Woodcock, R. W. (2001). *Essentials of WJ III tests of achievement assessment.* New York: John Wiley & Sons, Inc.

Mather, N., & Woodcock, R. W. (2001a). *Woodcock-Johnson III tests of achievement: Examiner's manual*. Itasca, IL: Riverside Publishing.

Mather, N., & Woodcock, R. W. (2001b). *Woodcock-Johnson III tests of cognitive abilities: Examiner's manual.* Itasca, IL: Riverside Publishing.

McGrew, K. S., & Woodcock, R. W. (2001). Woodcock-Johnson III technical manual. Itasca, IL: Riverside Publishing.

National Early Literacy Panel [NELP]. (2008). Developing early literacy: Report of the National Early Literacy Panel. Retrieved from http://www.nifl.gov/publications/pdf/NELPReport09.pdf

National Reading Panel [NRP]. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nichd.nih.gov/publications/nrp/upload/smallbook_pdf.pdf

The Nation's Report Card: Reading 2007 State Snapshot Report. (2007). National Center forEducation Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/stt2007/2007497NY4.pdf

New York State Report Cards. (n.d.). New York state testing and accountability reporting tool. Retrieved from https://www.nystart.gov/publicweb/

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*(1), 3-26. doi: 10.1207/s1532799xssr0801_2

Schneider, D., & Schneider, S. (2008). *Program assessments.* Amherst, NY: Genesee Hearing Services, LLP/Magic Penny Reading.

Schneider, D., & Schneider, S. (2009a). *Evidence-based early literacy results: Magic PennyReading* [Brochure]. Amherst, NY: Genesee Hearing Services, LLP/Magic Penny Reading.

Schneider, D., & Schneider, S. (2009b). *Why does the Magic Penny program work so well?* Retrieved from http://magicpennyreading.org/a/why-does-the-magic-penny-program-work-so-well

Schneider, D., & Schneider, S. (n.d.). *Magic Penny Early Literacy: Literacy development solutions* [Brochure].Amherst, NY: Genesee Hearing Services, LLP/Magic Penny Reading.

Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascolo, J. T. (2002). *Essentials of WJ III cognitive abilities assessment.* New York, NY: John Wiley & Sons, Inc.

Schuele, C. M., & Boudreau, D. (2008). Phonological awareness intervention: Beyond thebasics. *Language, Speech, and Hearing Services in Schools, 39*, 3-20. doi: 10.1044/0161-1461(2008/002)

Sciarrino, M. (2009). *My prayers have been answered.* Retrieved from http://magicpennyreading.org/a/letter-from-sr-sciarrino

Senechal, M. (2006). *The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3*. Washington, DC: National Institute for Literacy.

Shanahan, T. (2006). The shift from polarization in reading: Relying on research rather than compromise. In K. A. Dougherty Stahl & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 76-82). New York, NY: Guilford Press.

Solsken, J. W. (1993). *Literacy, gender, and work in families and in school.* Norwood, NJ: Ablex Publishing Corporation.

Stahl, S. A. (2006). Understanding shifts in reading and its instruction. In K. A. Dougherty Stahl & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 45-75). New York, NY: Guilford Press.

Stahl, S. A., Duffy-Hester, A. M., & Dougherty Stahl, K. A. (2006). Everything you wanted to know about phonics (but were afraid to ask). In K. A. Dougherty Stahl & M. C. McKenna (Eds.), *Reading research at work: Foundations of effective practice* (pp. 126-154). New York, NY: Guilford Press.

Trochim, W. M. K. (2006). Statistical power. *Research Methods Knowledge Base.* Retrieved from http://www.socialresearchmethods.net/kb/power.php

Trochim, W.M.K., & Donnelly, J.P. (2007). *The research methods knowledge base.* Mason, OH: Thomson.

U.S. Department of Education, Office of Elementary and Secondary Education. (2002). *Guidance for the reading first program.* Retrieved from www.ed.gov/programs/readingfirst/guidance.doc

Walpole, S., & McKenna, M. C. (2007). *Differentiated reading instruction: Strategies for the primary grades.* New York, NY: Guilford Press.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001a). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b). *Woodcock-Johnson III Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside Publishing.