School Psychology Quarterly
2011, Vol. 26, No. 1, 85–96

1045-3830/11/$12.00 DOI: 10.1037/a0022661

Early Number Skills: Examining the Effects of Class-Wide Interventions on Kindergarten Performance

Robin S. Codding
University of Massachusetts-Boston

Lisa Chan-Iannetta
Quincy Public Schools, Quincy, MA

Shauna George and Kristine Ferreira
University of Massachusetts-Boston

Robert Volpe
Northeastern University

The purpose of this study was to use multilevel modeling to compare the effects of KPALS alone and combined with goal setting and reinforcement to a control condition on early numeracy performance of 96 kindergarteners. Demographic variables were examined as moderators. Results differed according to early numeracy measure, with both versions of KPALS outperforming the control group for number identification, while KPALS plus goal setting and reinforcement improved performance on missing number and the TEMA-3. Demographic variables, but not treatment, accounted for performance on a measure of quantity discrimination. Posttest differences were found for TEMA-3 and NI, but no differences persisted at 4 months follow-up.

Keywords: mathematics, interventions, early numeracy, class-wide, kindergarten

Although recent legislation such as No Child Left Behind has placed emphasis on improving children's reading skills by using empirically supported reading interventions, 19% of fourth graders and 29% of eighth graders are performing below basic levels in mathematics (National Assessment of Educational Progress, 2009). These numbers are higher for urban districts representing 28% and 40% of fourth and eighth graders, respectively. This statistic is alarming considering students' failure in earlier mathematics material may place students at risk in higher level mathematics (Gersten & Chard, 1999), and, with curriculum reform, students are required to master more challenging material (Riley, 1997). In a review of six longitudinal sets of data, Duncan and colleagues (2007) found that the best predictors of later achievement were school-entry mathematics, reading, and attention, with mathematics skills serving as the strongest indicator. Early-mathematics research has illustrated that kindergarten children should understand and be able to operate on two principals: verbal counting and quantity discrimination (Okamoto & Case, 1996). By age 6, these underlying concepts should be coordinated into a mental number line representation, which is critical for future mathematics development (Moss & Case, 1999). Deficits in these and related areas, often described as number sense, are associated with later mathematics difficulties, and children living in low socioeconomic environments often arrive at school without the necessary prerequisites to develop these basic early skills (Griffin & Case, 1997b; Griffin, Case, & Siegler, 1994). These findings highlight the need for early mathematics prevention research to provide school professionals with empirically supported ideas so that students' difficulties can be effectively addressed.

A promising supplemental intervention administered class-wide may improve the early number skills of kindergarten children from low socioeconomic areas and across abilities. The

Robin S. Codding, Department of Counseling & School Psychology, University of Massachusetts-Boston; Lisa Chan-Iannetta, Quincy Public Schools, Quincy, MA; Shauna George and Kristine Ferreira, Department of Counseling & School Psychology, University of Massachusetts-Boston; and Robert Volpe, Northeastern University.

Funds for this project were made possible by the Joseph P. Healey Research Grant Program that was awarded to the first author.

Correspondence concerning this article should be addressed to Robin S. Codding, Department of Counseling & School Psychology, University of Massachusetts-Boston, 100 Morrissey Boulevard, Boston, MA 02125. E-mail: robin.codding@umb.edu

whole-class structure of this intervention is consistent with recommendations that empirically supported interventions be incorporated at primary (whole-class), secondary (small group), and tertiary (individual) levels (e.g., VanDerHeyden & Burns, 2005; VanDerHeyden, Witt, & Gilbertson, 2007). Furthermore, although kindergarten teachers may spend 25% of instructional time devoted to mathematics, activities tend to be integrated with other subjects rather than serving as the focus of teaching (Chung, 1994). Fuchs, Fuchs, and Karns (2001) developed Kindergarten Peer-Assisted Learning Strategies in Mathematics (KPALS), a downward extension of PALS-Math (Fuchs, Fuchs, Phillips, Hamlett, & Karns, 1995), which identifies activities that can be incorporated within typical classroom instruction for short periods of time. Peer-assisted learning strategies have considerable empirical support (Maheady, Mallette, & Harper, 2006), and a primary feature of the PALS series is class-wide peer tutoring that is monitored by teachers. KPALS was designed to target number knowledge by including quantity discrimination, number line, and general number recognition (e.g., written, symbolic, and pictorial representations of numbers) tasks. As an initial study on KPALS, Fuchs et al. (2001) randomly assigned 20 classrooms (168 children) to KPALS or control classrooms. Control classrooms consisted of typical basal curricular instruction, while KPALS classrooms replaced a portion of time devoted to mathematics instruction twice weekly to implementing KPALS activities. Results demonstrated that students receiving KPALS had higher rates of growth over time than typical classrooms regardless of whether initial number knowledge was low, medium, or disability status. However, effect sizes suggested that KPALS produced only small gains for high-performing students and moderate gains for all other achievement categories.

Although the results of this preliminary study are encouraging, a paucity of research has examined early prevention research in mathematics. Given the impact that early number knowledge has on later mathematics success, more research on interventions like KPALS is needed. We sought to enhance the previous research on KPALS in the following ways. First, although Fuchs et al. (2001) demonstrated improvements on a standardized, knowledge-based posttest mathematics-readiness measure, KPALS effects on fluency measures has not been examined. Gersten, Jordan, and Flojo (2005) suggested that a primary goal of early mathematics interventions is to increase accurate and efficient use of counting strategies and number combinations. These authors further propose that lack of proficiency with number skills is a key correlate of learning disabilities in mathematics. Therefore, examining whether interventions impact fluency of number skills is as important as improvements on standardized broad indicators of number knowledge (that primarily assess accuracy). Fortunately, technical adequacy of early number fluency measures has recently been demonstrated (e.g., Chard et al., 2005; Baglici, Codding, & Tryon, 2010). Specifically, this research found early numeracy fluency measures (i.e., Tests of Early Numeracy; TEN) that incorporate missing number, counting, quantity discrimination, and number identification tasks yielded adequate test-retest reliability and predictive validity on standardized outcome measures for kindergarten students. These measures may also potentially be useful for monitoring the progress of students participating in primary, secondary, and tertiary interventions (Clarke, Baker, Smolkowski, & Chard, 2008).

Second, the contribution of goal setting with reinforcement to KPALS has not been examined. In a meta-analytic review of 90 studies, Rohrbeck, Ginsburg-Block, Fantuzzo, and Miller (2003) found peer-assisted interventions that included interdependent reward contingencies and provided self-management opportunities (i.e., goal setting, reward selection and administration) had greater outcomes. Fuchs et al. (1997) specifically examined the impact of short, task-focused goal setting on math performance of students in Grades 2–4 who were receiving PALS. This research illustrated that pairing goal setting with PALS yielded more effort and more learning for low achieving students. Additionally, Morgan and Sideridis (2006) found in their meta-analysis that goal setting (with and without contingent reinforcement) led to the most growth in reading fluency over time compared to four other interventions. Together these findings suggest that if building fluency is an objective of early numeracy interventions, then goal setting and contingent reinforcement may be necessary components. Third, the maintenance of number skills following intervention termination has not been examined and is a critical for evaluating intervention effectiveness (American Psychological Association, 2002).

A number of demographic variables including age, gender, and primary spoken language may also influence the effects of early mathematics interventions. These are important considerations, as research suggests that academic achievement is related to economic and demographic variables, and all three elements contribute unique variance to disproportionate representation of monitory groups in special education (Hosp & Reschly, 2004). For example, age at kindergarten can be variable, and Jordan, Kaplan, Olah, and Locuniak (2006) demonstrated that older children performed better than younger children on number sense tasks, a difference maintained throughout the school year. Similarly, Aunio, Aubrey, Godfrey, Pan, and Liu (2008) found that, across cultures, older school-age children yielded better early numeracy performance than younger children. Impact of gender has mixed evidence, with some researchers finding similar early number skills among boys and girls (e.g., Dehaene, 1997), and others illustrating that either girls (Strand, 1999) or boys (Jordan et al., 2006) showed greater performance. Finally, English Language Learners (ELL) may yield poorer academic achievement than their English-speaking peers (Chang, 2008). Given the documented influence of these demographic and cultural variables on early number skills, it will be important to examine whether supplemental material offered within a class-wide intervention context is differentially beneficial to, for example, younger students who may begin the school year with fewer number sense skills or whether ELL students can make gains greater than their English-speaking peers.

Purpose of the Present Study

The purpose of this study was to extend the research on KPALS by comparing the isolated effects of KPALS with the combined effects of KPALS and goal setting with reinforcement to a control condition. To examine retention on early numeracy performance, we assessed fluency and general performance on a standardized measure 4 months following treatment termination. We also evaluated the impact of initial score on a standardized early numeracy measure in order to examine effects across initial skill as well as gender, age, and ELL status. Teacher treatment acceptability was examined to provide evidence of social validity (Eckert & Hintze, 2000). The following hypotheses were generated: (a) students receiving KPALS with goal setting and reinforcement (KPALS + GSR) would yield higher final scores and exhibit steeper slopes than students in the control or KPALS groups across measures; (b) students in the KPALS + GSR would yield higher posttest and maintenance performance on a standardized measure; (c) students in the KPALS + GSR group would demonstrate greater retention of skills than other students; and (d) initial skill level will impact treatment effectiveness. No hypotheses were generated for the demographic variables, as previous literature did not examine the impact of these factors on performance following preventative programming.

Method

Participants and Setting

Participants were recruited from two schools within one small urban district in the northeast region of the United States. Six kindergarten classrooms with a total of 96 students (class sizes ranged from 14 to 19) participated. Average school enrollment was 328.5 (range, 294 to 363), and the average percentage of low-income families was 58.6 (range, 41.3 to 70.4). Averaging across schools, 61.4% of students (range, 53.4 to 69.4) were Asian, 29% (range, 21.4 to 36.9) were Caucasian, 3.65% (range, 3.6 to 3.7) were African American, 2.8% (range, 2.4 to 3.3) were Latino, and 2.75% were Multi-Race (range, 2.7 to 2.8). An average of 59% of students' first language was not English (range, 51.8 to 66.3), and an average of 58.6% (range, 46.8 to 70.4) students had free or reduced lunch. We use ELL to represent the number of students for whom English was not the first language and these students include those that did and did not receive services.

Two students moved out of the district and were included in the growth modeling, but no data were available for follow-up. Fifty-four percent of the sample was male, and, for 43% of the sample, English was the second language.

The ethnic composition of the sample included 60% Asian, 33% Caucasian, 6% African American, and 1% Latino. Mean age across participants was 5 years, 5 months (range, 4 years, 8 months, to 6 years, 0 months) at the onset of the study.

The school district utilized the Houghton Mifflin Math (Houghton Mifflin Harcourt, 2005) curriculum series. Mathematics instruction ranged from 30 to 60 min either 4 or 5 days per wk. Daily instruction included group seatwork, which ranged from 20 to 30 min, and independent seatwork, ranging from 5 to 15 min across classes. Other instructional practices employed intermittently included small group (range, 10 to 15 min) and cooperative learning (range, 0 to 15 min).

Measures

Test of Early Mathematics Ability. The Test of Early Mathematics Ability, 3rd edition (TEMA-3; Ginsburg & Baroody, 2003), was individually administered three times (i.e., pre, post, and follow-up) in September, January, and May. The TEMA-3 is designed to assess mathematics knowledge for children ages 3 years to 8 years, 11 months, and includes items that address quantity discrimination, number literacy, and basic facts. The TEMA-3 has two parallel forms containing 72 items each, and an age-referenced standard composite score was computed for each participant. Test-retest (r = 82), alternate reliability (r = 97), and internal consistency (α = .94) are excellent and criterion validity (range, .54 to .91) is adequate. The TEMA-3 has been shown to be sensitive to student growth (Murphy, Mazzocco, Hanich, & Early, 2007). We used Form A to assess student performance at each testing session.

Tests of Early Numeracy. Tests of Early Numeracy (TEN; Clarke & Shinn, 2002) benchmark assessments were retrieved from AIMSweb (http://www.aimsweb.com). The TEN measures administered included number identification (NI), quantity discrimination (QD), and missing number (MN). One probe from each measure was administered three times, in September, January, and May. All probes were administered for one min and total items correct (TC) comprised each student's score. NI requires that participants orally identify numbers 0 to 10 when provided with a page of 56 numbers. Concurrent and predictive validity (range, .33 to .65) and alternate-form reliability (range, .91 to .92) are adequate (Clarke et al., 2008; Lembke & Foegen, 2009). MN requires that students orally supply the number omitted from a 3-component number line when provided with a page containing 21 number lines. Concurrent and predictive validity (range, .37 to .70) and alternate-form reliability (range, .59 to .72) are adequate (Clarke et al., 2008; Lembke & Foegen, 2009). QD requires students to select the larger number when presented with a worksheet containing 28 boxes of two numbers. Concurrent and predictive validity (range, .35 to .71) and alternate-form reliability (range, .83 to .89) are adequate (Clarke et al., 2008; Lembke & Foegen, 2009).

Progress monitoring. TEN progress monitoring measures obtained from AIMSweb (http://www.aimsweb.com) were individually administered on six occasions across 12 weeks (every other week) of treatment implementation. TEN measures consisted of NI, QD, and MN. Probes were identical in arrangement to benchmarks but represented alternate forms. In order to allow for growth, each probe contained two worksheets, thereby increasing the number of items possible to answer and decreasing the likelihood that prorated scores would need to be calculated.

Treatment acceptability. To assess teachers' perceptions, the Intervention Rating Profile-15 (IRP-15; Witt & Martens, 1983) was administered. The IRP-15 requires teachers to rate their level of agreement with 15 statements on a Likert-scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). The established reliability of the IRP-15 ranges from .88 to .98 (Martens, Witt, Elliott, & Darveaux, 1985), and, for our sample, was .74. Total scores on the IRP-15 range from 15 to 90, with higher scores indicating greater acceptance. An adapted version of the IRP-15 specifically referencing the treatment employed with teachers' students was distributed by the first author. Teachers who received the KPALS + GS intervention evaluated this combined intervention rather than each component. Teachers completed this form anonymously and returned the completed form in an envelope to the main office of each school.

Procedures

Teachers' classes were randomly assigned within schools to treatment groups. The number of students represented in each condition was as follows: control ($n = 35$), KPALS ($n = 31$), KPALS + GSR ($n = 30$). Three researchers, two school psychology-specialist-level graduate students, and the first author, trained in TEN and KPALS procedures, implemented interventions twice weekly for 20 min over 12 weeks. One 2-hr training session was conducted with the researchers, which included modeling, feedback, and practice. All assessment sessions were standardized and administered individually every other week on the second treatment day in desks outside the students' classrooms. Four additional specialist level graduate students were trained to administer TEN and TEMA assessments in addition to the researchers. All students had one or two semesters of graduate-level assessment (cognitive, academic, and CBM) training prior to the study implementation. Specific training sessions occurred across two 2-hr sessions that included modeling, feedback, and practice. Students needed to administer and score each instrument with 95% accuracy prior to the study's inception.

Kindergarten Peer Assisted Learning Strategies-Math. Kindergarten Peer Assisted Learning Strategies-Math (KPALS; Fuchs, Fuchs, Yazdian, Powell, & Karns, n.d.) is a supplemental class-wide peer tutoring program that includes 10 min of large group instruction and 10 min of peer-tutoring activities that span four areas. Twelve of the 16 total lessons spanning 3 of 4 areas were administered twice weekly across 12 consecutive weeks. The full 16 lessons were not implemented due to school constraints. Weeks were allocated to the following (corresponding to the sequence prescribed in the manual): 5 weeks, number concepts; 4 weeks, quantity comparisons (1 w/number lines); 2 weeks adding/subtracting concepts; and 1 week, training with students on KPALS rules and procedures.

KPALS + Goal Setting. Students in the KPALS + Goal Setting (KPALS + GSR) group received KPALS as described above. These students were also presented with a coloring worksheet containing vertical bars (one for each TEN measure) illustrating the number of problems previously completed correctly once every other week prior to conducting progress monitoring. Students colored in spaces on the worksheet representing their performance, were provided with new goals (15% increase over the previous scores; Carson & Eckert, 2003), and earned an educationally relevant prize (e.g., marker, eraser) if two of three goals (across measures) were achieved.

Control. Students in the control condition were assessed on TEN probes every other week on the same schedule and using procedures and probes identical to those administered to the other participants.

Interscorer Agreement and Procedural Integrity

Interscorer agreement was assessed for TC across 50% of assessment sessions for 25% of students. Comparisons between the researcher and independent scorer were conducted on a digit-by-digit basis. Mean percent agreement for TC was 98.6% (range, 96% to 100%), 99% (range, 99%), and 99% (range, 99% to 100%) for NI, MN, and QD, respectively. Procedural integrity was assessed via direct observation by an independent observer during 83% of the sessions across treatment groups and was 96% (range 83% to 100%). The observation form consisted of all the steps that were included in the KPALS scripted lesson plans, with a checkbox to indicate whether these steps were followed as written or not.

Results

Descriptive Statistics

No significant pretest differences across groups were found for age, $F(2, 93) = 0.98$, $p > .05$; proportions of male and female participants, $\chi^2(2, N = 96) = 0.22$, $p > .05$,; ELL, $\chi^2(2, N = 95) = 0.56$, $p > .05$; or ethnicity, $\chi^2(6, N = 92) = 5.92$, $p > .05$. Four separate ANOVAs were conducted with TEMA scores, NI, MN, and QD to examine for pretest differences among conditions. To control for the possibility of a Type I error, an alpha level of .0125 (.05/4) was used to determine significance. No significant differences were found (see Table 1).

Table 1
Pre-, Post-, and Follow-Up Data Across Treatment Groups

	Control			KPALS				KPALS + GSR			
Measures	*n*	*M*	*SD*	*n*	*M*	*SD*	d^a	*n*	*M*	*SD*	d^b
Pre-Test											
NI	35	37.00	22.92	31	34.48	17.97	—	29	39.52	21.03	—
MN	35	8.46	7.11	31	9.13	7.22	—	29	9.52	8.17	—
QD	35	15.97	10.70	31	17.81	7.63	—	29	16.69	10.48	—
TEMA-3	35	95.63	17.06	31	94.32	13.51	—	29	97.41	20.43	—
Post-Test											
NI	35	49.97	10.72	30	57.63	17.61	0.52	29	60.27	23.61	0.56
MN	35	15.22	8.66	30	15.73	7.58	0.06	29	18.76	10.60	0.37
QD	35	26.00	11.84	30	30.33	12.84	0.35	29	29.00	14.82	0.22
TEMA-3	35	100.31	10.92	30	103.50	11.17	0.29	29	106.24	16.94	0.42
Follow-Up											
NI	34	59.41	20.54	30	61.63	17.23	0.12	28	67.68	23.46	0.37
MN	34	18.76	8.13	30	18.67	7.80	−0.01	28	21.07	10.30	0.25
QD	34	32.59	13.61	30	32.20	12.64	0.30	28	34.11	14.67	0.11
TEMA-3	34	101.00	12.00	30	103.93	10.12	0.26	28	104.32	16.87	0.22

[a] Represents the effect size for KPALS compared to the control group. [b] Represents the effect size for KPALS + GSR compared to the control group. KPALS = Kindergarten Peer Assisted Learning Strategies; KPALS + GSR = Kindergarten Peer Assisted Learning Strategies with Goal Setting and Reinforcement; NI = number identification; MN = missing number; QD = quantity discrimination; TEMA-3 = Test of Early Mathematics Ability, 3rd Edition.

Intervention Effectiveness

SAS PROC MIXED was used for the analyses due to the longitudinal nature of the data and our interest in monitoring individual progress of each student. The ML estimation method and unrestricted covariance structure were employed. ML estimations are the most common, as they are asymptotically unbiased, normally distributed, and efficient (Singer & Willett, 2003). The unrestricted covariance structure was the best fit after comparing the different variance-covariance structures. A 3-level model was used with Class as Level 3, Students as Level 2, and Repeated Measures as Level 1. Data from the 6 treatment sessions and the pretest data were analyzed for each TEN measure. Ninety-five percent ($n = 91$) of our sample had complete data across all assessments, and a total of 26 observation points were missing.

Predictors were group membership (KPALS, KPALS + GSR, Control), initial TEMA-3 scores, ELL status, age, and gender. Group membership was reference coded (1 to 3) so that variables representing group membership could be used to predict differences in students' final TC and growth rates. Scores from the control group served as the reference with which the effects of the analyses in the final model were compared. Initial score on the TEMA-3 was also coded to represent students performing at or above (1) and those below (2) the sample average. ELL status was determined by students receiving ELL services. Age was dummy-coded to represent students at or above (1) and those below (2) the sample average age in months.

The unconditional model was fitted to determine whether students' TC varied over time across measures. For NI, the time parameter showed that, on average, students were gaining 2.91 TC from session to session. The average final score (i.e., the last day of treatment) was 58.68 TC (see Table 2). Interclass correlations (ICC) were .24 and .79 for levels 2 and 3, respectively, meaning that 24% of NI variation occurred across students and 74% across classes. For MN, average gains were 1.19 TC. Average final score was 17.66 TC (see Table 3). ICC coefficients were .19 and .83 for levels 2 and 3, respectively. For QD, students gained 1.70 TC and the average final score was 28.80 TC (see Table 4). ICC coefficients were .30 and .81 for levels 2 and 3, respectively.

The final model included the following predictors for both the intercept (final performance within the treatment phase) and slope: (a) initial TEMA-3 score, (b) group membership, (c) ELL

Table 2
Multilevel Prediction Models for Number Identification

Parameters	Predictors	Unconditional model	Final model
Fixed effects			
Final status	Intercept	58.68 (2.80)	50.07 (3.82)
	KPALS + GSR		13.27 (3.92)**
	KPALS		9.13 (3.92)**
	TEMA-3		11.26 (3.32)**
	ELL		−11.37 (3.25)**
Growth	Intercept	2.91 (0.01)**	−3.37 (0.35)**
	KPALS + GSR		−1.64 (0.35)**
	KPALS		−1.17 (0.29)**
	At/above mean TEMA-3		1.74 (0.29)**
	At/above mean age		0.69 (0.27)*
Random effects	Level 1	84.73 (4.7)**	76.85 (4.3)**
	Final status (student)	26.19 (27.2)	0.00
	Final status (class)	317.61 (49.1)**	210.63 (32.1)**
Goodness-of-fit	Deviance statistics	5796.60	5563.50

Note. Parenthetical information represents standard errors. KPALS = Kindergarten Peer Assisted Learning Strategies; KPALS + GSR = Kindergarten Peer Assisted Learning Strategies with Goal Setting and Reinforcement; TEMA-3 = Test of Early Mathematics Ability, 3rd Edition; ELL = English Language Learners.
* $p < .05$. ** $p < .001$.

status, (d) age, and (e) gender. For NI final performance, KPALS and KPALS + GSR were significant predictors. Students in both treatment groups had higher final scores than students in the control group. For MN final performance, only students receiving KPALS + GSR had higher final scores than students in the control group. For QD final performance, treatment group was not a significant predictor. For NI, MN, and QD, pre-Test TEMA-3 scores and ELL status were significant predictors. Students who scored at or above average initially on theTEMA-3 had higher final scores than those who performed below average, and ELL stu-

Table 3
Multilevel Prediction Models for Missing Number

Parameters	Predictors	Unconditional model	Final model
Fixed effects			
Final status	Intercept	17.66 (1.1)	15.22 (1.8)
	KPALS + GSR		3.92 (1.8)*
	KPALS		1.06 (1.9)
	TEMA-3		4.64 (1.5)*
	ELL		−3.96 (1.5)*
Growth	Intercept	1.19 (0.0)**	−1.66 (0.14)**
	KPALS + GSR		−0.26 (0.1)*
	KPALS		−0.04 (0.1)
	At/above mean TEMA-3		0.29 (0.1)*
	At/above mean age		0.39 (0.1)**
	Male		0.34 (0.1)**
Random effects	Level 1	11.68 (0.6)**	11.07 (0.6)**
	Final status (student)	2.67 (3.8)	0.00
	Final status (class)	58.92 (9.0)**	48.06 (7.2)**
Goodness-of-fit	Deviance statistics	4239.20	4167.90

Note. Parenthetical information represents standard errors. KPALS = Kindergarten Peer Assisted Learning Strategies; KPALS + GSR = Kindergarten Peer Assisted Learning Strategies with Goal Setting and Reinforcement; TEMA-3 = Test of Early Mathematics Ability, 3rd Edition; ELL = English Language Learners.
* $p < .05$. ** $p < .001$.

Table 4
Multilevel Prediction Models for Quantity Discrimination

Parameters	Predictors	Unconditional model	Final model
Fixed effects			
Final status	Intercept	28.80 (1.8)	28.15 (3.0)
	KPALS + GSR		3.92 (2.9)
	KPALS		2.99 (3.0)
	TEMA-3		5.81 (2.1)*
	ELL		−7.52 (2.1)**
Growth	Intercept	1.79 (0.01)**	−2.42 (0.2)**
	KPALS + GSR		−0.19 (0.2)
	KPALS		−0.15 (0.2)
	At/above mean TEMA-3		0.70 (0.2)**
	At/above mean age		0.86 (0.2)**
Random effects	Level 1	26.50 (1.5)**	25.10 (1.4)**
	Final status (student)	11.28 (10.9)	2.80 (5.1)
	Final status (class)	114.29 (17.5)**	84.01 (13.1)**
Goodness-of-fit	Deviance statistics	4931.10	4756.20

Note. Parenthetical information represents standard errors. KPALS = Kindergarten Peer Assisted Learning Strategies; KPALS + GSR = Kindergarten Peer Assisted Learning Strategies with Goal Setting and Reinforcement; TEMA-3 = Test of Early Mathematics Ability, 3rd Edition; ELL = English Language Learners.
* $p < .05$. ** $p < .001$.

dents performed less well than their English-speaking peers.

For NI slope, KPALS and KPALS + GSR were significant predictors. Thus, students in both treatment groups had greater growth than the control group from session to session. For MN slope, KPALS + GSR was a significant predictor with these students yielding higher rates of session growth. For QD slope, group was not a significant predictor. For NI, MN, and QD, TEMA-3 scores and age were significant predictors. Students who were at or above average on the TEMA-3 and/or were older in age had lower rates of growth than those below average. For MN, gender was a significant predictor and males had lower rates of growth than females.

Deviance statistics demonstrated that the final model was a better fit than the unconditional model for NI, $\chi^2(12) = 233$, $p < .001$; MN, $\chi^2(12) = 71.3$, $p < .001$; and QD, $\chi^2(12) = 174.9$, $p < .001$, respectively (Singer & Willett, 2003). Psuedo-R^2 ([unconditional random effect—final model random effect/the unconditional random effect] $\times$ 100) was used to estimate a global effect size for each level of analysis across measures (Singer & Willett, 2003). For NI, 9% of the variance was accounted for at Level 1, all of the variance was accounted for by the predictors at the student level, and 34% of the variance was accounted for by the predictors at the class level. For MN, 5% of the variance was accounted for at Level 1, all of the variance was accounted for at the student level, and 18% of the variance was accounted for at the class level. For QD, 5% of the variance was accounted for at Level 1, 75% at the student level, and 26% at the class level. It should be noted that there is not a consensus regarding effect-size calculation for MLM analyses, and these effect sizes are not comparable to traditional effect sizes such as Cohen's *d* (Peugh, 2010).

Power analyses were conducted by first estimating the slope and the standard deviation according to the growth curve model and then using PROC POWER. These analyses yielded power of 0.81 for NI; however, power was less than 50% for MN and QD.

Posttest and Maintenance

To examine posttest (one month posttreatment) and skill maintenance 4 months following intervention termination, between-group ANCOVAs were conducted for TEMA-3 and TEN (see Table 1). After adjusting for preintervention TEMA-3 scores and class, there was a significant difference between groups at posttest, $F(2, 94) = 4.20$, $p = .02$, $\eta_p = .086$. Post hoc analyses illustrated that the mean score for KPALS + GSR was significantly different from controls. No differences were found at 4 months follow-up, $F(2, 92) = 1.38$, $p = .26$, $\eta_p = .03$. Pretest scores on TEN and class served as the covariates for the corresponding analyses. A significant difference was found between groups at posttest for NI, $F(2, 89) = 3.25$, $p = .04$, $\eta_p = .068$. No significant differences were yielded for maintenance.

Unweighted effect sizes were calculated using Cohen's *d* for each dependent measure assessed across groups for posttest and follow-up data. For posttest data, effect sizes ranged from small to medium when comparing control group scores to both KPALS and KPALS + GSR. For follow-up data, effect sizes ranged from negligible to small when comparing the control group to both experimental groups.

Acceptability

Mean scores on the IRP-15 measure for the intervention groups were acceptable for KPALS ($M = 78.5$) and KPALS + GSR ($M = 89.0$). Scores higher than 52.5 indicate acceptable ratings.

Discussion

The purpose of the present study was to extend the research on KPALS by comparing its effects with a treatment package that added goal setting and reinforcement on early numeracy performance to a control condition. This study extended the extant literature by examining (a) the impact of these conditions on fluency measures as well as a standardized measure, (b) retention, and (c) demographic variables. Our results partially supported the hypotheses by finding that KPALS + GSR produced significantly higher scores at the end of the intervention, and greater growth between sessions, than the control group for the NI and MN measures. KPALS + GSR students had significantly higher NI and TEMA-3 scores at posttest compared to the control group; however, maintenance of group performance differences after 4 months was not observed. Demographic variables were significant predictors of each fluency measure.

Treatment affected both the NI and MN performance but not QD performance. That is, final status and growth over time were significantly lower for the control group as compared to both KPALS groups for NI. The strong impact of both KPALS groups on NI corresponds with the specific focus of this curriculum on numbers (Fuchs et al., 2001). For MN, final status and growth was significantly higher for the KPALS + GSR group compared to the controls. It is possible that, had the sample size been larger, KPALS also would have yielded significant differences from the control group. The impact of treatment group on TEMA-3 performance illustrated that immediately following treatment termination, only students in the KPALS + GSR performed better than the control group. Positive effects of goal-setting on fluency building have been noted elsewhere (e.g., Codding, Chan-Iannetta, Palmer, & Lukito, 2009; Fuchs et al., 1997; Morgan & Sideridis, 2006) and are consistent with the notion that providing reinforcement may encourage students to persist on tasks that can be performed accurately (Daly et al., 2007). Goal setting also helps focus student performance on specific target skills, perhaps increasing active learning (Morgan & Sideridis, 2006). That is, by providing reinforcement and feedback on performance toward academic goals, goal setting encourages students' effort toward their own learning (Fuchs et al., 1997; National Mathematics Advisory Panel, 2008).

The failure to find treatment effects on the QD measure may have occurred due to its restricted range. Some research has shown that growth is observed within kindergarten on MN but not QD measures (Baglici et al., 2010), raising the question of whether QD can be used to monitor treatment progress, while other research has suggested that QD measures growth adequately (Clarke et al., 2008; Lembke & Foegen, 2009). Interestingly, the KPALS curriculum closely aligns to the quantity discrimination measure and provides more examples of this type of activity than those represented by missing number. Four lessons focused specifically on the concepts of more or less, with one formally introducing number lines. However, Griffin and Case (1997a) suggested that number identification and counting skills, which might be represented by the missing number measure, are skills that develop earlier than quantity comparisons. Important to note is that the between-student variance was higher for QD than the other measures, and our data illustrate that ELL status, age, and initial TEMA-3 scores moderated performance.

Across all fluency measures, pre-Test TEMA-3 scores predicted final status and growth, with students initially falling in the average or above average range yielding higher performance than below average students; however, these students made less growth over time than students below average. This finding is consistent with Fuchs et al. (2001), who found stronger effect sizes for students with disabilities and those with low and medium achievement status as compared to students with high achievement status. It is also possible that scores reached a ceiling, so progress could not be demonstrated. Unfortunately, these data also suggest that below-average students did not "catch up" to their higher-performing peers despite making more growth.

Consistent with research suggesting that achievement is impacted by demographic variables (Hosp & Reschly, 2004), ELL status and age served as significant predictors for all fluency measures. ELL students demonstrated lower initial and final scores than non-ELL students, but status did not impact growth over time, suggesting that gains made by all students were similar. Research by Griffin and Case (1997a) and Chang (2008) has suggested that language barriers can impede students' ability to demonstrate number knowledge skills within the classroom. This finding may suggest that ELL students would benefit from more targeted instruction that focuses on the language associated with early number concepts. It is also possible that the early numeracy measures were not culturally or linguistically appropriate for the students in our sample. Conversely, age did not affect final TC, meaning that students performed similarly regardless of age. However; students who were younger than the mean of the sample yielded significantly more growth over time, which differs from findings by Jordan et al. (2006), who demonstrated that age differences persisted through the academic year. This difference may be due to the fact that Jordan and colleagues (2006) were examining growth over time for students receiving the core curriculum without any additional intervention components. Gender was only a predictor for MN, with males exhibiting significantly less growth from session to session than females. That gender did not predict final performance likely corresponds with higher male scores during pretest than female scores (albeit these differences were not significant), potentially suggesting that males reached a ceiling before their female counterparts.

Although students receiving KPALS and KPALS + GSR performed better than the control group 4 months following the end of treatment, these differences were not significant. This may have resulted from the study's limitations. It is possible that the findings herein would be different with a larger sample size, as we did not have enough power to detect significant differences for two of our measures (i.e., MN and QD). In addition the entire 16-week KPALS program was not implemented; however, the omitted area consisted of adding and subtracting numbers, which was not the primary interest of the investigation. It could also be that having researchers, rather than teachers, implement the program impacted the generalization of these findings. Teacher implementation might have facilitated reinforcement of KPALS lessons throughout the day and may have led to better coordination between the intervention and core curriculum. However, in that case, contamination among teachers could have occurred. Although still possible in our study, the researchers retained the materials following each session, so it was unlikely that teachers instituted practices from the other classes. Related, a core curriculum or common method of instruction was not present among the six teachers. Although it is unknown to what extent variation in instructional content and time impacted the results, we randomly assigned classrooms to conditions and controlled for classroom in our analyses. It is possible that the positive results associated with the TEMA-3 scores is the result of practice effects, as we used the same form for pre-, post- and follow-up test, but this is not likely given that these assessments were administered approximately 16 weeks apart. Students for whom English is not a first language may have performed more poorly because the assessments were provided in English rather than in their primary language or both languages. Although we assessed teacher acceptability of the intervention conditions, we did not seek input from the students, which is an important component of acceptability and should be collected in future research. Finally, standard error estimates were high, and Christ (2006) suggested that slope is more meaningful when stable and with small standard errors of estimate.

Future research should take into account instructional differences in the classrooms by also controlling for, or evaluating, the type and total amount of instructional time allocated to mathematics. Alternatively, it would be useful for researchers to spend an equivalent amount of time in control classrooms in order to further equate the conditions. Although this study examined demographic moderators, such as ELL, associated with the treatment gains, future research should specifically examine the impact of other intervention packages on the mathematics performance of ELL students. It might be important for these studies to consider the level of language proficiency rather than just ELL status. Future research should extend this study with larger sample sizes. The utility of TEN and other early fluency measures for monitoring progress over time is worthy of continued investigation. In particular, the QD measure as a progress-monitoring tool should be further examined, as the data herein suggested that demographic markers, rather than the treatment, accounted for performance variance. Finally, future research might develop, examine, and compare various approaches to improving number skill development in kindergarten.

References

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57,* 1052–1059.

Aunio, P., Aubrey, C., Godfrey, R., Pan, Y., & Liu, Y. (2008). Children's early numeracy in England, Finland, and People 's Republic of China. *International Journal of Early Years Education, 16,* 203–221.

Baglici, S. P., Codding, R. S., & Tryon, G. (2010). Extending the research on tests of early numeracy: Longitudinal analyses over two years. *Assessment for Effective Intervention, 35,* 89–102.

Carson, P. M., & Eckert, T. L. (2003). An experimental analysis of mathematics instructional components: Examining the effects of student-selected versus empirically-selected interventions. *Journal of Behavioral Education, 12,* 35–54.

Chang, M. (2008). Teacher instructional practices and language minority students: A longitudinal model. *The Journal of Educational Research, 102,* 83–97.

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30,* 3–14.

Christ, T. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35,* 128–133.

Chung, D. H. (1994). *Young children's acquisition of mathematical knowledge and mathematics education in kindergarten* (Unpublished doctoral dissertation). Iowa State University, Ames, IA.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29,* 46–57.

Clarke, B., & Shinn, M. R. (2002). *Tests of early numeracy measures (TEN): Administration and scoring of AIMSweb early numeracy measures for use with AIMSweb*. Eden Prairie, MN: Edformation.

Codding, R. S., Chan-Iannetta, L., Palmer, M., & Lukito, G. (2009). Examining a class-wide application of cover-copy-compare with and without goal setting to enhance mathematics fluency. *School Psychology Quarterly, 24,* 173–185.

Daly, E. J., III, Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36,* 562–581.

Dehaene, S. (I1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Eckert, T. L., & Hintze, J. M. (2000). Behavioral conceptions and application of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly, 15,* 123–148.

Fuchs, L., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995). Acquisition and transfer effects of class wide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24,* 604–620.

Fuchs, L., Fuchs, D., Yazdian, L., Powell, S., & Karns, K. (n.d.). *Peer assisted learning strategies: Kindergarten math teacher's manual*. Nashville, TN: Center on Accelerating Student Learning.

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergarteners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101,* 495–585.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goal on low achieving students with and without learning disabilities. *American Educational Research Journal, 34,* 513–543.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33,* 18–28.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38,* 93–304.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early numeracy mathematics ability* (3rd ed.). Austin, TX: Pro-Ed.

Griffin, S. A., & Case, R. (1997a). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3,* 1–49.

Griffin, S. A., & Case, R. (1997b). Wrapping up: Using peer commentaries to enhance models of mathematics teaching and learning. *Issues in Education, 3,* 115–134.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for the first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children, 70,* 185–199.

Houghton Mifflin Harcourt. (2005). *Houghton Mifflin Math.* Boston, MA: Author.

Jordan, N., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77,* 153–175.

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first grade students. *Learning Disabilities Research & Practice, 24,* 12–20.

Maheady, L., Mallette, B., & Harper, G. F. (2006). Four classwide peer tutoring models: Similarities, differences, and implications for research and practice. *Reading & Writing Quarterly, 22,* 65–89.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. (1985). Teacher judgments concerning the acceptability of school based interventions. *Professional Psychology: Research and Practice, 16,* 191–198.

Morgan, P., & Sideridis, G. D. (2006). Contrasting the effectiveness of fluency interventions for students with or at-risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. *Learning Disabilities Research & Practice, 21,* 191–210.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education, 302,* 122–147.

Murphy, M. M., Mazzocco, M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD.

National Assessment of Educational Progress. (2009). *Mathematics assessment*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monologues of the Society for Research in Child Development, 61,* 27–59.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85–112.

Riley, R. W. (1997). In math and college-going, middle school makes all the difference. *Middle School Journal, 29,* 3–7.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning intervention with elementary school students. A meta-analytic review. *Journal of Educational Psychology, 95,* 240–257.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Strand, S. (1999). Ethnic group, sex, and economic disadvantage: Associations with pupils' education progress from baseline to the end of key stage 1. *British Educational Research Journal, 25,* 179–202.

VanDerHeyden, A. M., & Burns, M. K. (2005). Examination of various measure of mathematics proficiency. *Assessment for Effective Intervention, 33,* 215–224.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model identification of children for special education. *Journal of School Psychology, 45,* 225–256.

Witt, J. C., & Martens, B. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20,* 510–517.