Targeting the Specific Vocabulary Needs of At-Risk Preschoolers: A Randomized Study of the Effectiveness of an Educator-Implemented Intervention

Delphine Vuattoux · Christa Japel · Eric Dion ·
Véronique Dupéré

Abstract This randomized study examined the effectiveness of a preschool stimulation program created to teach words that had been selected by considering the needs of the target population of children. Twenty-two educators and their group of at-risk preschoolers ($N=222$, $M_{age}=4.27$ years) were assigned to one of two conditions: control or intervention. In the latter condition, educators had to read specifically developed storybooks to their group and conduct stimulation activities. Despite the training and support they received, educators implemented the intervention with varying degrees of fidelity. Nonetheless, intent-to-treat comparison of the two conditions indicates that children in the intervention condition learned the meaning of a much greater number of words than their peers in the control condition. In addition, efficacy subset analyses that took into account fidelity of implementation show that the greatest gains were made by children who had an educator who had implemented the intervention reliably. Strategies for scaling up the intervention and optimizing its implementation are discussed.

Keywords Preschoolers · Vocabulary · Explicit instruction · Preschool educators · Implementation fidelity

D. Vuattoux · C. Japel · E. Dion
Département d'éducation et formation spécialisées, Université du Québec à Montréal, Montreal, QC, Canada

V. Dupéré
École de psychoéducation, Université de Montréal, Montreal, QC, Canada

C. Japel (✉)
DEFS, UQAM, C. P. 8888, Succ. Centre-Ville, Montreal, QC, Canada H3C 3P8
e-mail: japel.christa@uqam.ca

As a key marker of language and cognitive development, oral vocabulary directly reflects the quality of stimulation young children receive (Hoff 2013). Unfortunately, many children learn only a limited number of words, and these children tend to come from low-income families (Ryan et al. 2006). Their limited vocabulary is likely due to a lack of verbal stimulation at home (Hart and Risley 1995) or their parents' limited proficiency in the majority language (August and Shanahan 2006). Independent of its specific causes, the lag in vocabulary development can be substantial and have lasting consequences. Findings suggest the existence of a 2,000-word gap between children with more restricted vocabulary and those with more developed vocabulary (Biemiller and Slonim 2001); and being on the wrong side of this gap is predictive of learning problems, especially reading difficulties (e.g., Dickinson and Porche 2011). To a lesser extent, a restricted vocabulary has also been associated with dysfunctional peer relationships, perhaps because adequate verbal skills are necessary to understand and negotiate emotion-arousing situations (Menting et al. 2011; Nelson et al. 2011). Stimulating at-risk preschoolers' vocabulary development could help these children adequately adjust to school demands.

Limitations of the family environment can, in theory, be compensated by interventions in other settings, notably high-quality programming in preschool or childcare centers (Dupéré et al. 2010; Japel 2008). Indeed, preschool environments are historically important foci of prevention efforts (Aber et al. 2012). However, although stimulating language development is a stated goal of preschool programs (Neuman and Dwyer 2009), educators working in these settings rarely provide adequate stimulation to children (Japel et al. 2005; Justice et al. 2007). This contradiction between institutional goals and actual practices could be a result of the paucity of well specified and adapted interventions.

Preventive interventions that aim at modifying existing practices—rather than advocating completely new ones—

are probably more likely to be accepted by practitioners (Miller and Shinn 2005). Fortunately, there is a common practice in preschool settings that has potential for vocabulary development: storybook reading. Its potential lies in the fact that a richer vocabulary is found in storybooks than in everyday interactions (Cunningham and Stanovich 1998). Reading to children thus exposes them to unfamiliar words and can help them enlarge their vocabulary, at least when it is done correctly. As shown by Hargrave and Sénéchal (2000), for instance, children must be actively engaged in the reading routine for benefits to occur. In this study, at-risk children were assigned to one of two conditions. In the first, reading was carried out without soliciting their participation. In the second, educators stimulated children's participation and interest by asking them question on the story and encouraging them to also ask questions. Children's participation was associated with greater vocabulary gains, but these gains were modest. These results suggest that participation is necessary, but not sufficient, to ensure learning.

This is not surprising considering how difficult it is for children to learn the meaning of a word (for a detailed descriptive study, see, e.g., Johnson and Mervis 1994). It is thus necessary to teach word meanings systematically and explicitly (Marulis and Neuman 2010). For instance, Coyne et al. (2007) asked kindergarten teachers to read to their group and to explicitly teach the meaning of a random selection of difficult words found in the storybook. A simple definition of the meaning of these words was presented on multiple occasions, children were asked to notice the words during the reading of the storybook, and they were encouraged to use these words and explore their meaning. For comparison purposes, other difficult words were only defined in passing or not taught at all. Learning was minimal, except when words had been taught explicitly. Furthermore, even with explicit instruction, children remembered the meaning of only half of the words (see also Coyne et al. 2004).

These findings highlight the need to improve the effectiveness of explicit instruction. The content of instruction could also be more relevant. Interventions typically use commercially available storybooks and instruction is centred, by necessity, on the difficult words found in the books. Unsurprisingly, the resulting selection is interesting but somewhat arbitrary (e.g., "sushi," "zookeeper"; Coyne et al. 2004). If the goal is indeed to help children learn to read when they eventually enter school (e.g., Justice et al. 2007), instruction has to focus on words that are both useful for reading and unknown to the children from the community or population segment targeted by the intervention (for the importance of adapting prevention programs, see Barrera et al. 2011). In addition, because each instructed word is found in only one storybook (e.g., Hargrave and Sénéchal 2000), once the book is read, the child is not exposed to the word again. This is problematic since

reviewing word meanings could improve retention. Using a specifically developed set of storybooks allows not only a control over the selection of words but also the inclusion of multiple reviews.

The Current Study

The goal of this study was to examine the effect of an educator-implemented stimulation program on the vocabulary of at-risk 4-year-olds from predominantly low-income communities. The originality of the intervention lies in the use of specifically developed storybooks designed to present selected words on multiple occasions. The words had been selected by considering the needs of the target child population. Educators in community-based childcare centers were assigned to a control (regular activities) condition or to an intervention condition. In the latter condition, educators were asked to implement the intervention with their group. We expected that children in the intervention groups would learn a greater number of word meanings than their peers in the control condition.

Method

Participants

Twenty-two educators (95.5 % women) participated in this study conducted in Montreal (Quebec, Canada) and surrounding areas. They worked in 12 francophone non-profit childcare centers implementing the provincial play-based preschool curriculum (Hohmann et al. 2008). These centers were located in urban ($n=4$) or semi-rural ($n=8$) low-income communities. All educators agreed to be randomly assigned, with their entire group, to one of the two conditions. Educators in the two conditions were similar. They had a mean of 12.1 years (SD=8.2) of experience. The majority (77.3 %) held a postsecondary vocational degree in childcare, and only one of them was trained as a teacher. Their sole material incentive to participate was the possibility of keeping the intervention material (which was also offered to educators in the control condition at the end of the study).

The sample included 222 children (46.8 % girls, $M_{age}=$ 4.27 years) for whom parental consent had been obtained (82.2 % of children in the groups). Thirty-six percent of the participating children were from a family with an income below the poverty line, 37.6 % of the children were from a family in which the mother did not have a postsecondary education, and 32.6 % of the children spoke a language other than French at home. No between-condition differences were observed on these variables.

Because of absences or family relocations over the course of the study, data are missing for 10.4 % ($n=23$) and 17.1 % ($n=38$) of the children at pre- and posttest, respectively. In order to avoid introducing biases and facilitate multilevel analyses, multiple imputations were used to handle missing data (Allison 2001). To take into account imputation uncertainty, multiply imputed data sets were created (through the SAS MI procedure) and used for all analyses. Ten data sets were created, the maximum allowed by the Hierarchical Linear Models software (Raudenbush et al. 2004). The analyses were performed separately on each of the ten imputed data set, and the parameter estimates obtained in each set were recombined, resulting in a final single set of estimates.

Design

The 22 participating groups of 4-year-old children were in 12 childcare centers, with one to four groups per center ($M=$ 1.83 group). The groups (of children with their educator) were randomly assigned to the two conditions, irrespective of center, with the only constraint being that a slightly greater number of groups be assigned to the intervention condition ($n=12$) than to the control condition ($n=10$). Since it was groups that were assigned to the conditions, most centers with more than one group (five out of eight) had at least one group in each condition. The decision to assign a greater number of educators to the intervention condition reflected our concern that some of these educators could decide to discontinue their participation. As it turned out, the 22 educators completed the study.

Intervention

Word Selection To select word meanings for instruction, we used the only available list of words used orally in francophone first-grade classrooms as a starting point (Préfontaine and Préfontaine 1968). This list was updated by removing anachronisms (e.g., "smoking pipe") and by adding contemporary words (e.g., "computer") found in the Ministry of Education's approved first-grade reading material. We also removed words that are known by all 4-year-olds (e.g., "dog," "banana") according to standardized tests (Dunn et al. 1993; Trudeau et al. 1999).

The resulting 1,133-word list was examined by eight first-grade teachers working with either low- or middle-income students (these teachers participated in an unrelated study). The two groups of teachers were asked to identify, among the 1,133 words, those that were both useful for reading and unlikely to be familiar to their students. There was a consensus among the two groups of teachers that 799 of the 1,133 words were likely to be familiar to both low- and middle-income students. These words were discarded as

too easy. As it turned out, a majority of teachers (75 % or more) in the two groups considered the remaining 334 words to be useful for reading and likely to be known by middle-income students but not by their low-income peers. Further exclusion of articles, plural forms, and verb conjugations resulted in a shortened list of 273 words.

The shortened version of the list was presented to 32 educators working in childcare centers located in low- or middle-income communities. The two groups of educators (who did not participate in the experiment) had to indicate whether each word was known by a majority of 4-year-olds in their group. According to their answers, approximately 80 % of the words on the shortened list were known only by middle-income children. Most words on the shortened list were thus judged by both educators and first-grade teachers to present a special challenge for children of low-income communities. Because stories needed to be coherent and interesting, we were able to include only 175 of the words (71 nouns, 56 verbs, 44 adjectives, and 4 adverbs) of the shortened list in the storybooks.

Instructional Activities An award-winning author of children's books was consulted to ensure that the style and content of the 30 storybooks developed for the intervention was appropriate. Each of the storybooks presents a narrative story and is accompanied by a pedagogical guide that includes a detailed lesson plan and a definition for all instructed words (see Coyne et al. 2004). Each page of the storybooks contains a short text with an original illustration depicting the content of the text and emphasizing, when applicable, the meaning of the instructed (magic) word in its context (Fig. 1). Five or six words are taught per storybook, and four to six previously taught words are integrated in the story. On average, each word was reviewed in 2.4 storybooks. Each book was read twice, during two consecutive sessions. Half of the words were taught during the first session, and the other half was taught during the second session. Educators were asked to conduct four 15-min sessions per week. Educators read all of the storybooks to their group, in the same prescribed order.

Before reading the storybook, the educator reminded the children that they had to raise their hand upon hearing a magic word. The educator then presented the magic words (e.g., "striped," "defend," "immense"), asked the children to repeat these words, and presented a short (six word or less) and simple definition of their meaning (e.g., that "striped" means "has lines of different colors"; Beck et al. 2002). When a magic word was encountered while reading the storybook, the educator repeated its definition, asked children to pronounce the word, and pointed to the relevant elements of the illustration (e.g., "a striped bee").

After the reading, a recall activity was conducted. The educator showed the page of the storybook on which the

Mimi is a little bee just like all the others. She has black and yellow stripes. You can't recognize her because she has the same stripes as all the other bees. She is a curious and cheerful bee.

Fig. 1 Translated sample of instruction material: the page of a storybook illustrating the meaning of an instructed word (stripe)

meaning of each word was illustrated, repeated the definition, and asked the children to identify the word (e.g., "Do you remember the word for a thing with lines of different colors?"). This was followed by two verbally mediated activities that encouraged the children to use the words and to reflect on their meaning. The first was a discrimination activity. The educator asked the children to make the distinction between contexts in which the word could or could not be applied (e.g., "Is a tiger striped?" "What about a white piece of paper?"). The second was a generalization activity. Children were asked to use the word in a context different from the one in which it had been taught (e.g., "Today, Lea wears a shirt with lines of different colors. Which magic word does this make you think of?"). Children were free to look at the storybooks when the session was over.

Instruments

General Receptive Vocabulary This aspect of vocabulary knowledge was assessed at pretest with the French version of the Peabody Picture Vocabulary Test (PPVT; Dunn et al. 1993). A series of pages with four pictures was presented to

the child who was asked to identify which picture corresponds to the word presented by the research assistant. A standardized score was calculated based on the number of correct answers given by the child. The score for the original version of the instrument is strongly correlated with cognitive abilities and achievement scores (Dunn et al. 1993). Although the PPVT score is a valid index of general vocabulary, it is not sensitive to intervention effects (Coyne et al. 2004), probably because the words used for the assessment do not correspond to those taught during the intervention.

Specific Expressive and Receptive Vocabulary As recommended by the National Reading Panel (National Institute of Child Health and Human Development 2000), learning of instructed (specific) word meanings was assessed with researcher-developed measures whose content is aligned with the intervention. For the pre- and posttest expressive assessment, 15 words were randomly selected from the pool of instructed words, equally sampling nouns, verbs, and adjectives. Following Biemiller and Slonim's (2001) procedure, a sentence was composed for each of the selected words (e.g., "Matthew has a *striped* ball."). The research assistant read the sentence and asked the child, "What does [the word] mean?" Children could verbally explain the meaning of the word or, in some cases, mime it (e.g., spread arms for "gigantic"). The answer was noted and later categorized as complete (1 point) when the instructed definition was restated (e.g., "very beautiful" for "magnificent"), when the important aspects of the meaning were mentioned (e.g., "house" and "bird" for "nest"), or when a synonym was given (e.g., "giant" for "enormous"). It was categorized as incomplete (0.5 point) when the child mentioned only one important aspect of the meaning (e.g., "brothers" for "twins"). Finally, it was categorized as incorrect (0 point) when the child offered an erroneous definition (e.g., "joke" for "join") or no answer. A total score was computed by summing up the individual item scores (α=.87). Thirty-three percent of the assessments were coded by a second research assistant, which yielded an inter-rater agreement of 97 %.

In addition to the expressive assessment conducted at pre- and posttest, a receptive assessment was conducted at posttest. This last assessment is easier than the expressive one because the child is not required to articulate the meaning of the word (Coyne et al. 2007). The receptive assessment resembles the PPVT: A series of four pictures was shown to children, who had to identify the picture corresponding to the word presented by the research assistant. Receptive knowledge of 14 of the 15 words used for the expressive assessment was assessed. The total score corresponds to the number of correctly identified pictures (α=.74).

Fidelity of Implementation The fidelity with which educators and children in the intervention condition implemented

the instructional activities was formally observed by a research assistant during a predetermined session at mid-intervention. The researcher-developed, 61-item checklist was adapted from similar instruments (e.g., Fuchs et al. 1997). With young children involved, obtaining a high score is considered possible only when the intervention has been correctly implemented on a regular basis for a significant period of time, that is, when the intervention has become a well-rehearsed routine. The checklist items describe expected behaviors on the part of the educators (e.g., "Presents the magic words") and children (e.g., "Raise their hand when hearing a magic word") during each activity: storybook reading (26 items), recall (13 items), discrimination (9 items), and generalization (9 items). Four items also refered to the participation and attention of the whole group of children (e.g., "The majority of children were attentive"). Each behavior is scored as observed or not, and the percentage of observed behaviors is calculated. A second research assistant was present for 25 % of the observations. Interrater agreement on the items of the checklist was 99 %.

Procedure

Assessments Children were assessed individually, in September (pretests) and in February (posttests), in a quiet space close to the group's room by an undergraduate or graduate student in education or psychology. The research assistants ($n=8$) made sure to motivate the children and make them feel comfortable before conducting the assessments. Because being unfamiliar with the examiner may have a negative impact on the performance of young at-risk children (Fuchs and Fuchs 1986), each child was assessed by only one research assistant. This research assistant also observed implementation and was thus not blind to the conditions of the study. In order to avoid introducing biases, research assistants were trained to conduct assessments following detailed protocols.

Training and Support At the beginning of October, educators assigned to the intervention condition learned how to implement the intervention during a half-day workshop, either individually or in groups of two or three. Each activity was explained in detail, demonstrated by the trainer, and practiced by the educators. Commonly occurring problems were discussed and examples of lesson plans contained in the manual (Japel et al. 2008) were reviewed. Educators also received a calendar for implementing the intervention.

Educators were asked to implement the intervention four times per week from October to February. During this period, each group was visited twice by the assistant responsible for conducting the pre- and posttest assessments. The goal of these visits was to help the educator to reliably implement the intervention through encouragement, advice, or direct support. During one of the visits, fidelity of implementation was observed using the checklist and formal feedback was given to the educator based on these observations. Assistants knew the intervention well and had received directives on how to give constructive feedback to the educators. Additional in-class support was available upon request. Throughout the project, the research team met twice a month to discuss issues pertaining to assessment and intervention implementation.

Results

Fidelity of Implementation Despite the efforts made to ensure that instructional activities would be uniformly well implemented in the groups assigned to the intervention condition, the overall fidelity of implementation varied from 21.3 % to 100 % ($M=66.2$ %, SD=30.5). Considered for each activity separately, fidelity was, on average, acceptable for storybook reading ($M=75.9$, SD=21.8), but poorer for recall ($M=65.1$, SD=37.6), discrimination ($M=60.9$, SD=46.0), and generalization ($M=54.9$, SD=42.4). Children's participation and attention were also less than optimal ($M=50.0$, SD=45.2), suggesting that some educators had problems managing their group.

To take into account the disparity in fidelity of implementation, intent-to-treat and efficacy subset analyses were conducted (Lachin 2000). Intent-to-treat analyses consider only the assigned condition, irrespective of the imperfect nature of intervention implementation. This type of analysis avoids introducing a bias in favor of the intervention, a bias that could occur, for instance, if educators implemented the instructional activities with greater fidelity in the more receptive groups (i.e., the groups who would have made the most progress in any case). For their part, efficacy subset analyses are used to estimate how the effectiveness of the intervention varies as a function of the fidelity with which it is implemented. Because groups have not been randomly assigned to the different levels of implementation, these are correlational analyses and their results must be interpreted accordingly.

Condition Equivalence at Pretest Since instructional activities were offered to groups rather than individuals, children's vocabulary scores were not statistically independent. This dependency was taken into account in the multilevel analyses. Two-level models, with individual scores (level 1) nested in groups (level 2) were estimated. For pretest comparisons and intent-to-treat analyses, the experimental condition was treated as a group (level 2) dichotomous variable (0=control, 1=intervention). Accordingly, group means were compared rather than individual scores.

Table 1 presents results as a function of condition. At pretest, children in the two conditions had a comparable

general receptive vocabulary, $B=2.24$, $t(df=20)=0.68$, n.s. Furthermore, they had similarly low specific expressive vocabulary scores. Indeed, there was a floor effect on these scores: At pretest, children knew, on average, the meaning of approximately three words, and about 16 % of them did not know the meaning of any words. This is interesting from a substantive point of view because it suggests that the words selected for instruction were of an appropriate level of difficulty. Statistically, however, the departure from normality precludes a conventional analysis. To circumvent the problem, we conducted two complementary analyses: one for count (Poisson) distributions and another for dichotomous (Bernoulli) outcomes (0=no word meaning, 1=one or more word meaning). No between-condition differences were observed, both for the count distribution, $B=0.35$, $t(df=20)=1.87$, n.s., and the dichotomous outcome, OR=1.31, 95 % CI=0.41–4.22, n.s.

Intent to Treat These analyses were conducted on the posttest specific vocabulary scores. We began by examining the percentage of children with a score of zero in the two conditions. Although the number of children who did not know any word meanings in the intervention condition was three times less than in the control condition, the difference did not reach significance, OR=3.03, 95 % CI=0.64–14.25, n.s. Even if it was less pronounced than at pretest, a floor effect continued to be observed for expressive vocabulary scores, especially in the control condition. In a reverse manner, there was a ceiling effect on the receptive vocabulary scores (i.e., many children had the maximum score), mostly in the intervention condition. In order to derive a normally distributed score, we standardized the correlated ($r=.78$) specific expressive and receptive vocabulary scores separately and computed the average of the two scores for each child. An analysis for normally distributed outcomes indicates that intervention groups had, on average, higher composite scores than control groups, $B=0.69$, $t(df=20)=3.86$, $p<.001$. To calculate an effect size (see Hedges 2007), the value of the HLM regression parameter associated with the (dichotomous) condition variable (0.69) was divided by the standard deviation of group means (SD=0.45). A very strong effect size (ES=1.53) was observed.

Efficacy Subset In this section, effectiveness is examined as a function of the fidelity of intervention implementation by the educator, a group (level 2) variable. It is assumed that the intervention was not implemented at all in the control condition groups, that is, where fidelity is equal to 0 %. To detect nonlinear effects, the squared value of the fidelity score was also introduced in the model. Since no such effects were detected, only analyses with the raw fidelity scores are presented.

At pretest, there is no relation between fidelity and general vocabulary, $B=0.04$, $t(df=20)=1.11$, n.s., or specific receptive vocabulary, $B=0.01$, $t(df=20)=1.71$, n.s. In other words, it does not appear that the fidelity with which the educator implemented the intervention was influenced by the group's initial language skills. However, a positive relation is observed between fidelity and the posttest composite specific vocabulary score, $B=0.008$, $t(df=20)=3.71$, $p<.01$. To examine the strength of this relation in terms of effect size, we used the approach proposed by Stoolmiller et al. (2000). Group means on the composite vocabulary score were estimated for nil (0 %, control condition), low (30 %), good (60 %), and excellent (90 %) implementation (Table 2). Each of these last three means were compared to the mean of the control condition groups by dividing the arithmetic difference between the two means by the pooled standard deviation of the group means (SD=0.45). A moderate effect size was observed for a low implementation (ES=0.56), a strong one for a good implementation (ES=1.13), and a very strong one for an excellent implementation (ES=1.69).

Discussion

This study aimed at examining the effectiveness of an explicit vocabulary intervention created to address the specific needs of a target population of at-risk preschoolers. Intent-to-treat analyses indicate that the intervention was highly effective, despite a suboptimal implementation by some educators, whereas efficacy subset analyses suggests that a more reliable implementation was associated with greater vocabulary gains. In what follows, we consider factors that could be responsible for the intervention effectiveness, but also try to explain why some educators failed to implement this intervention correctly.

Intent-to-treat analyses provide a conservative estimate of the effectiveness of unevenly implemented interventions (Lachin 2000). Nevertheless, comparison of the two conditions, irrespective of fidelity of implementation, indicates that the intervention had a very large effect (Cohen 1988) on vocabulary as measured by the posttest composite score. This effectiveness can be attributed, in part, to the relative absence of systematic language stimulation in the control condition groups. In other words, we probably compared our intervention to regular practices that, unfortunately, did not provide much competition (for a similar assessment of preschool practices, see Justice et al. 2007; Neuman and Dwyer 2009). As our findings show, however, many educators are able to stimulate language development. In previous studies, preschool vocabulary interventions implemented by preschool educators had an almost negligible effect (Marulis and Neuman 2010). The problem apparently lies with the tested interventions rather than with the educators.

Weak regular practices cannot be the only factor explaining the effectiveness of our intervention. It is

Table 1 Vocabulary knowledge by condition irrespective of fidelity of implementation (intent-to-treat analyses)

Time administered/ Measure	Control		Intervention		SD[g]
	% [e]	Mean [f]	% [e]	Mean [f]	
Pretest					
General receptive[a]		93.3		95.5	5.1
Specific expressive[b]	18.0	2.7	14.8	3.6	1.0
Posttest					
Specific expressive[b]	12.3	4.7	4.4	7.7	1.5
Specific receptive[c]		7.8		9.8	0.8
Composite specific[d]		−0.3		0.3	0.3

[a] Standardized score on the French adaptation of the Peabody Picture Vocabulary Test (Dunn et al. 1993)

[b] Number of correctly defined words (range, 0–15)

[c] Number of correctly identified pictures (range, 0–14)

[d] Average of the standardized scores on the specific expressive and receptive measures

[e] Proportion of children with a score of zero

[f] Estimated mean of group means in each condition

[g] Estimated SD of group means pooled for the two conditions

probably effective, in an absolute sense, because of three of its distinctive characteristics. First, in contrast to other studies on vocabulary instruction conducted with preschoolers (e.g., Hargrave and Sénéchal 2000), it relies on an explicit form of instruction. Our study shows that this age group, like older children (Marulis and Neuman 2010), respond well to clear explanations and carefully planned, high-interest vocabulary activities. In a sense, this is not surprising since positive results have also been observed for preschool programs relying on explicit instruction to teach

Table 2 Vocabulary knowledge by level of fidelity of implementation (efficacy subset analyses)

Time administered / measure	Degree of implementation				SD[d]
	Nil	Poor	Good	Excellent	
Pretest					
General receptive[a]	94.5	95.9	97.2	98.6	4.8
Specific expressive[b]	3.2	3.6	4.0	4.3	1.0
Posttest					
Composite specific[c]	−0.3	−0.0	0.2	0.5	0.4

Estimated means of group means

[a] Standardized score on the French adaptation of the Peabody Picture Vocabulary Test (Dunn et al. 1993)

[b] Number of correctly defined words (range, 0–15)

[c] Average of the standardized scores on the specific expressive and receptive measures

[d] Estimated SD of group means pooled for the two conditions

phonological awareness, another key precursor of reading (for a review, see Phillips et al. 2008). Second, in our intervention words are frequently reviewed, and this could have improved retention and lead to better outcomes at posttest. Children are not very good at remembering the meaning of words in the absence of an extended exposition (e.g., Johnson and Mervis 1994). Finally, in contrast with what is usually done, the instructed words were selected based on an empirical identification of the needs of the targeted population. The close correspondence between the curriculum and the children's needs, made possible by the use of specifically developed storybooks, could have facilitated learning.

Of course, adapting an intervention to the specific needs of children from a given community or population segment raises the possibility that the intervention will be less relevant for other children. Preventive interventions can indeed be effective with one population segment, but not with another (Barrera et al. 2011). How is it possible, then, to identify the groups of children for which the intervention is relevant and should be "scaled up"? Conversely, how can scientists or practitioners determine if and when an intervention should be adapted in order to adequately meet the needs of other groups? Castro et al. (2004) argue that such decisions should be guided by empirically based strategies, strategies like the procedure used in this study to select words for instruction. Before scaling up the intervention to a potentially distinct population segment, this inexpensive procedure could assist in determining whether the selection of words is relevant.

Our results are not entirely positive. Approximately one educator out of four did not implement the intervention correctly, and correlational efficacy subset analyses suggest that the uneven fidelity of implementation had consequences for learning. Unsurprisingly, the largest effect size was observed for groups with an excellent fidelity of implementation. In a more surprising manner, however, a non-negligible effect size was also apparent for groups with a poor implementation. In these groups, educators apparently read the storybooks correctly, but neglected to explicitly teach word meanings. Being read to by an adult can, by itself, be beneficial for children with an initially larger vocabulary (Sénéchal et al. 1995), and the progress made by this minority of children could explain the non-negligible effect size. The other children, those with an initially more restricted vocabulary, may not have made sufficient progress.

Why is it that some educators were not able, or did not see the need, to correctly implement the intervention despite the training and on-site support they received? Changing one's practices is probably difficult for some educators given the conflicting messages they receive concerning their role. The official play-based curriculum encourages educators to limit themselves to observe and to sporadically guide

children's play, and it thus can make any systematic interventions appear superfluous or worse. In this context, some educators may not see the need to overcome the obstacles they meet when implementing an intervention. Perhaps, they would be more motivated to persevere if they were given regular feedback on the progress made by their group as a result of their implementation of the intervention (Landry et al. 2009).

The long-term goal of early vocabulary instruction is to help prevent reading problems and their cascading negative consequences. Experiencing reading problems early on places students at high risk of low school achievement, grade retention, and referral for special education, difficulties that, in addition to their emotional toll, ultimately increase the odds of dropping out of high school (Alexander et al. 2001; Dauber et al. 1993; Hibel et al. 2010). In theory, participation in a preschool vocabulary program should lower the risk of reading problems when children begin elementary school. In practice, it remains to be seen whether vocabulary gains are large and persistent enough to improve readiness to learn to read. Although there are encouraging findings (e.g., Loftus et al. 2010), only a handful of intervention studies have conducted a longitudinal follow up of significant duration. Long-term intervention studies must be undertaken in order to determine the amount of early vocabulary stimulation required to effectively reduce the risk of reading problems.

In spite of its strengths, our study is thus limited by the absence of a follow up. Other limitations must be noted. First, we did not assess factors that could have contributed to the variations in fidelity of implementation (e.g., personal characteristics of the educators), limiting our capacity to understand the role of these factors. Second, the procedure used to select words for instruction was based on the opinion of educators and teachers rather than on a direct assessment of children's needs (e.g., Biemiller and Slonim 2001). Our procedure has the merit of being inexpensive, but its validity has to be firmly established. Third, fidelity of implementation was observed formally on only one occasion. It would probably have been preferable to conduct multiple observations. Nonetheless, the strong relation between vocabulary gains and our fidelity of implementation score suggests that this score was reliable. Finally, we should have collected evidence confirming that our sample of educators only minimally stimulates language development in the context of their normal practices (Japel et al. 2005). We can conclude in affirming, however, that many educators can stimulate at-risk children's language development when provided with the right prevention program.

References

Aber, L., Morris, P., & Raver, C. (2012). Children, families and poverty: Definitions, trends, emerging science and implications for policy. *Social Policy Report, 26*, 1–28.

Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teacher College Record, 103*, 760–822. doi:10.1097/00004583-199007000-00014.

Allison, P. D. (2001). *Missing data*. Thousand Oaks: Sage.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah: Lawrence Erlbaum.

Barrera, M., Jr., Castro, F. P., & Steiker, L. K. H. (2011). A critical analysis of approaches to the development of preventive interventions for subcultural groups. *American Journal of Community Psychology, 48*, 439–454. doi:10.1007/s10464-010-9422-x.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*, 498–520. doi:10.1037/0022-0663.93.3.498.

Castro, F. P., Barrera, M., Jr., & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*, 41–45.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.

Coyne, M., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*, 145–162. doi:10.1207/s15327035ex1203_3.

Coyne, M., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*, 74–88. doi:10.2307/30035543.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*, 1–8.

Dauber, S. L., Alexander, K. L., & Entwisle, D. R. (1993). Characteristics of retainees and early precursors of retention in grades: Who is held back? *Merrill-Palmer Quarterly, 39*, 326–343.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82*, 870–886. doi:10.1111/j.1467-8624.2011.01576.x.

Dunn, L. M., Theriault-Whalen, C. M., & Dunn, L. M. (1993). *Adaptation française du peabody picture vocabulary test – revisée [The French adaptation of the peabody vocabulary scale - revised]*. Toronto: Psycan.

Dupéré, V., Leventhal, T., Crosnoe, R., & Dion, E. (2010). Understanding the positive role of neighborhood socioeconomic advantage in achievement: The contribution of the home, child care and school environments. *Developmental Psychology, 46*, 1227–1244. doi:10.1037/a0020211.

Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56*, 243–262. doi:10.2307/1170377.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive

to diversity. *American Educational Research Journal, 34*, 174–206. doi:10.2307/1163346.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*, 75–95. doi:10.1016/S0885-2006(99)00038-1.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H Brookes.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370. doi:10.3102/1076998606298043.

Hibel, J., Farkas, G., & Morgan, P. L. (2010). Who is placed into special education? *Sociology of Education, 83*, 312–332. doi:10.1177/0038040710383518.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*, 4–14. doi:10.1037/a0027238.

Hohmann, M., Weikart, D., & Epstein, A. S. (2008). *Educating young children* (3rd ed.). Ypsilanti: HighScope Press.

Japel, C. (2008). Risques, vulnérabilité et adaptation: Les enfants à risque au Québec [Risk, vulnerability, and adaptation: At-risk children in Quebec]. *Choix IRPP, 14*, 1–46. Retrieved from http://www.irpp.org/fr/choices/archive/vol14no8.pdf.

Japel, C., Vuattoux, D., Dion, E., & Bergeron, L. (2008). *Mimi et ses amis. Activités d'enseignement du vocabulaire au préscolaire* [Mimi and her Friends : Vocabulary Activities for Preschoolers]. Unpublished document. Université du Québec à Montréal. Montréal, Qc.

Japel, C., Tremblay, R. E., & Côté, S. (2005). *Quality counts! assessing the quality of daycare services based on the Quebec longitudinal study of child development. Choices, 11, 5*. Montreal: Institute for Research on Public Policy. (published Dec. 13th, 2005) Retrieved from http://www.irpp.org/fr/choices/archive/vol11no4.pdf.

Johnson, K. E., & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology, 30*, 418–435. doi:10.1037/0012-1649.30.3.418.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2007). Quality of language and litteracy instructions in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 23*, 51–68. doi:10.1016/j.ecresq.2007.09.004.

Lachin, J. M. (2000). Statistical considerations in the intention-to-treat principle. *Controlled Clinical Trials, 21*, 167–189. doi:10.1016/S0197-2456(00)00046-5.

Landry, S., Anthony, J., Swank, P., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*, 448–465. doi:10.1037/a0013842.

Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language

and literacy difficulties. *Learning Disabilities Research and Practice, 25*, 124–136. doi:10.1111/j.1540-5826.2010.00310.x.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary training on word learning: A meta-analysis. *Review of Educational Research, 80*, 300–335. doi:10.3102/0034654310377087.

Menting, B., van Lier, P. A. C., & Koot, H. M. (2011). Language skills, peer rejection, and the development of externalizing behavior from kindergarten to fourth grade. *Journal of Child Psychology and Psychiatry, 52*, 72–79. doi:10.1111/j.1469-7610.2010.02279.x.

Miller, R. L., & Shinn, M. (2005). Learning from communities: Overcoming difficulties in dissemination of prevention and promotion efforts. *American Journal of Community Psychology, 35*, 169–183. doi:10.1007/s10464-005-3395-1.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00–4754). Washington: US Government Printing Office.

Nelson, K. E., Welsh, J. A., Vance Trup, E. M., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Language, 31*, 164–194. doi:10.1177/0142723710391887.

Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-K. *The Reading Teacher, 62*, 384–392. doi:10.1598/RT.62.5.2.

Phillips, B. M., Clancy-Menchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children. *Topics in Early Childhood Special Education, 28*, 3–17. doi:10.1177/0271121407313813.

Préfontaine, R. R., & Préfontaine, G. (1968). *Vocabulaire oral des enfants de 5 à 8 ans au Canada français [Oral vocabulary of five-to eight-years old French-Canadians]*. Montréal: Beauchemin.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & duToit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwoood: Scientific Software International.

Ryan, R. M., Fauth, R. C., & Brooks-Gunn, J. (2006). Childhood poverty: Implications for school readiness and early childhood education. In B. Spodek & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed., pp. 323–346). Mahwah: Lawrence Erlbaum.

Sénéchal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87*, 218–229. doi:10.1037/0022-0663.87.2.218.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trail targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology, 68*, 296–306. doi:10.1037/0022-006X.68.2.296.

Trudeau, N., Frank, I., & Poulin-Dubois, D. (1999). Une adaptation en français Québécois du MacArthur Communicative Development Inventory [A French-Canadian adaptation of the MacArthur Communicative Development Inventory]. *La Revue D'Orthophonie Et D'Audiologie, 23*, 61–73.