



HAL
open science

Effects of policy and educational interventions intended to reduce difficulties in literacy skills in grade 1

Jean Ecalte, Christophe Gomes, Pauline Auphan, Laurent Cros, Annie Magnan

► **To cite this version:**

Jean Ecalte, Christophe Gomes, Pauline Auphan, Laurent Cros, Annie Magnan. Effects of policy and educational interventions intended to reduce difficulties in literacy skills in grade 1. *Studies in Educational Evaluation*, 2019, 61, pp.12 - 20. 10.1016/j.stueduc.2019.02.001 . hal-03484902

HAL Id: hal-03484902

<https://hal.science/hal-03484902v1>

Submitted on 20 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Effects of policy and educational interventions intended to reduce difficulties in literacy skills in Grade 1

Jean Ecalle^{1,2}, Christophe Gomes³, Pauline Auphan^{1,2,3}, Laurent Cros³, & Annie Magnan^{1,2}

¹ *Laboratoire EMC Etude des Mécanismes Cognitifs (EA 3082), Université Lyon2, France*

² *LabEx Cortex Lyon ANR-11-LABX-0042, France*

³ *Association Agir pour l'Ecole, Paris, France*

Corresponding author:

Jean Ecalle, Lab EMC - Lyon2 5, avenue Mendès-France 69676 BRON Cédex France

ecalle.jean@wanadoo.fr

Effects of policy and educational interventions intended to reduce difficulties in literacy skills in Grade 1

Abstract. In two large-scale studies, the effects of policy and educational interventions on literacy skills were examined in children schooled in zones with specific educational needs. To calculate the potential effects of such interventions, treatment-effects estimators with nearest neighbor matching were used. In Study 1 with policy intervention ($N=1095$), children in experimental group (Exp) were assigned to small classes (12 pupils) and others in control group (Cont) to normal-sized classes (20-25 pupils). At the end of Grade 1, the effect sizes in favor of Exp were .14 and .22 in word reading and spelling. In Study 2 with educational interventions ($N=2803$), children in Exp benefit from an evidence-based practice, i.e. a code-focused intervention (phonology, letter knowledge, decoding and fluency) developed by the *Association Agir pour l'Ecole* (Act for School) and conducted by teachers in small groups for children with low performance at the beginning of Grade 1. The effect sizes of interventions in various literacy skills were from .12 to .32. This set of results obtained in France is in accordance with those described in other countries. To conclude, a double intervention with small classes and targeted educational approaches could be one of the best ways of reducing inequalities during learning to read.

Highlights.

- Two large-scale studies were conducted in zones with specific educational needs.
- The reduction of the class size had an effect on reading and spelling.
- Targeted interventions in favor of low achievers had an effect on literacy skills.
- Effect sizes that vary from .12 to .32 are found in large samples.

The reduction of reading difficulties is an important political aim in all democratic countries. Over the last two decades, international surveys have provided information about students' reading levels in many countries and, in some cases, governments have taken measures to combat the difficulties observed in a proportion of students. We can distinguish between two types of intervention: those decided on by the ministry of education (policy interventions) and those which are implemented by teachers in their classes (educational interventions). After examining the work performed in these domains, two large-scale studies conducted in France will be presented. The aim is to put these two kinds of interventions in perspective and then discuss their limitations and advantages, in particular in terms of their financial cost.

State of the art: what we know about interventions on literacy skills

Policy interventions

The causal effect of class size on pupils' school performance is currently an important issue in the education field. Class size is considered as one important factor that can increase student achievement (Hattie, 2005). Although, the effects of reducing class size on student performance have been investigated on a number of occasions in the US, fewer studies have investigated class size effects in European countries. More particularly, few studies have examined the effects of class size in France (for a review, see Bressoux, 2016). Research on class size differences has mainly focused on the effects of such differences on academic achievement (Nye, Hedges, & Konstantopoulos, 1999; Nye, Hedges, & Konstantopoulos, 2002) and on teacher-pupil interactions (Blatchford, Basset, & Brown, 2011; Blatchford, Edmonds, & Martin, 2003; Blatchford, Moriarty, Edmonds, & Martin, 2002; Blatchford & Russell, in press a; b). Positive findings concerning class size effects on children's performance come mainly from the data reported by a large-scale experiment in Tennessee

(U.S.), named Project STAR (Student Teacher Achievement Ratio). This program ran from 1986 to 1989 and made major experimental studies in this field possible. In the STAR project, 7000 pupils were randomly assigned to small (13–17 students), regular (22–25) and regular with teaching assistance classes. Evidence from the STAR project shows that the benefits of class size reduction are most marked in the early stages of a child's schooling, i.e. kindergarten through Grade 3 (5-8 years), as well as in children from minority ethnic backgrounds (Finn & Achilles, 1999; Krueger, 1999). Moreover, the advantage of being in a small class in early grades has both immediate and long-lasting effects on academic achievement (Finn, Gerber, & Boyd-Zaharias, 2005). The long-term effect has not been fully established. Some authors consider that the effects are still present when pupils move to regular classes from grades 4 to 6 (Nye, Hedges, & Kostantopoulos, 2000), while others do not think that this is the case. In fact, a reanalysis of the large data set from the STAR study using multi-level modelling reduces the sizes of the effects (Goldstein & Blatchford, 1998) and Ding and Lehrer (2010), who made use of a structural model to account for non-random attrition, found a significant effect in kindergarten and first grade, but none in later grades. Moreover, the STAR program design has been criticized a/ for a failure to pre-test the participating pupils to ensure that the assignment was truly random across class sizes and b/ because some of the participating children were reassigned across class sizes due to behavioral problems or at the request of the parents.

Another program was conducted in North Carolina in 1991, partly in response to the criticisms of the STAR program. In this program, termed Success Starts Small, pupils were assigned to classes of either 15 or 25 students in first through third grade. Achilles, Harman and Egelson (1995) found that the children in the smaller classes achieved test scores that were .45 and .56 standard deviations above those of their peers in the larger classes on the mathematics and reading tests, respectively.

Blatchford, Goldstein, Martin, and Browne (2002) reported on the first-year UK longitudinal study of 9330 reception-year pupils from a group of randomly selected schools in England. They showed a clear effect of class size differences on academic attainment in literacy and mathematics, and an effect for low baseline achievers and the socio-economically disadvantaged. In addition, this study has been the subject of research about connections between class size and teaching interactions (Blatchford & Russell, in press a; b).

In sum, research has shown that reducing the class size can have an impact on classroom organization and teaching practices, student motivation and engagement, and sometimes also on school performance (Anderson, 2000; Konstantopoulos & Sun, 2014). The results also indicated that, although all pupils benefited from smaller classes, reductions in class size did not narrow the achievement gap in mathematics and reading between low and high achievers (Konstantopoulos, 2008).

Recently, Zyngier (2014) examined class size reduction and its effect on student achievement by analyzing 112 peer-reviewed studies from 1979-2014. He showed that the majority of these studies found that smaller classes in the first four years of school have a significant impact on student achievement and help narrow the achievement gap, especially for children from culturally, linguistically and economically disenfranchised communities. However, this is particularly true when smaller classes are combined with appropriate teacher pedagogies suited to reduced student numbers.

Finally, research on class size effects has, as we have seen, addressed relationships with academic outcomes and classroom processes respectively. Current research on class size is concerned with the pedagogical changes needed in order to make the most of class size reductions. There is evidence that teachers do not always change their teaching in small classes (Blatchford & Russell, in press a). The challenge is to propose innovative pedagogical interventions in the context of reduced-size classes.

Fewer studies have been specifically designed to research class size effects on reading skills. A large-scale class-size reduction experiment was undertaken in France during the 2002-03 school year. It involved 200 first-grade classes. In 100 experimental classes, the size was drastically reduced to 10-12 pupils and these were compared with 100 full-size control classes (21.3 pupils on average). Different literacy skills were assessed: word recognition, vocabulary, spelling, phonology, reading comprehension. A global literacy score was calculated. Over the two-year period, the children's literacy performance was assessed five times (beginning, middle and end of grade 1; beginning and end of grade 2). Bressoux (2016) reported a positive effect of reduced class size at the end of grade 1 (+.22) on a global literacy score but found no evidence of a long-term effect. Reading lessons were also observed, but only during first grade. Differences were noted in the way the teachers managed the classes: the teachers of the reduced classes had far fewer discipline or class management problems and they worked together with small groups in the class more frequently.

For the same study, Ecalle, Magnan, and Gibert (2006) focused on two basic literacy skills in grade 1, namely reading (word recognition task) and spelling (word production task). After controlling for initial literacy skills in the two groups at the beginning of first grade, they found that class size had a small impact on word reading and word spelling. Moreover, children in reduced classes were more engaged in their learning activities at the end of first grade (Bressoux & Lima, 2004). Overall, the results indicated a positive association between class size and pupil's performance. The results obtained in France for class size effects are consistent with those found in other countries.

More recently, Shen and Konstantopoulos (2017) examined class size effects on fourth graders' reading achievement in Europe using PIRLS (Progress in International Reading Literacy Study) data. They studied whether class size reduction impacts reading performance

in eight European countries (Bulgaria, Germany, Hungary, Italy, Lithuania, Romania, Slovakia, and Slovenia) over a 10-year period. Other studies have examined class size effects on reading achievement in Greece using PIRLS data (Konstantopoulos & Traynor, 2014). Overall, the results indicate that class size effects on reading achievement are not significant. In summary, the relationship between class size and achievement is not clear (Konstantopoulos, 2011). There is limited evidence about class size effects, in particular with regard to reading achievement (4th grade to 8th grade). The results depend on the design of the studies, the assessment methods and also on average class size and the size of the reduction. It appears that new studies of early grades are necessary. In the first study below, we reviewed the results of the class size reductions in France implemented in first grade by the Ministry of Education. This also represents an opportunity to put these results into perspective relative to those obtained in recent educational interventions.

Educational interventions

Teachers are faced with the challenge of teaching children with different academic levels in their classrooms. For example, reading abilities will vary in a first grade class made up of 20 children. One child may be reading at a kindergarten level, while another has not yet acquired letter knowledge. In such cases, the teacher must find ways to adapt lesson plans to meet the learning abilities of both pupils, while also accommodating the needs of the other 18 children in the class. One solution to this challenge is to implement differentiated instruction in the classroom. Differentiated instruction is a pedagogical approach that supports effective and specific instruction for all students (Hattie, 2009; 2012) and consequently helps promote egalitarian education. In a differentiated classroom, educators divide their time, resources, and efforts in a way that makes it possible to effectively teach children who have varying backgrounds, commitment and skill levels.

Reading difficulties are one of the main causes of academic failure in France. There are many reasons why children may not learn to read. Regardless of the cause, specific teaching and early intervention can improve children's learning outcomes and reduce the overall incidence of reading difficulties. Today, there is considerable evidence that reading difficulties can be remediated and even prevented through early intervention (Torgesen, 2002; Vellutino & Zhang, 2008). Researchers argue that early intervention - in kindergarten and Grades 1 and 2 - is more effective than later intervention because of the intensity and duration of the activities required if later intervention is to be effective and the difficulty of remediating fluency rates (Ehrhardt, Huntington, Molino, & Barbaresi, 2013; Foorman, Breier, & Fletcher, 2003). Wanzek and Vaughn (2007) found larger effects for intervention provided in Grades K-1 than 2-5. Based on scientific knowledge, it is possible to promote best practices to stimulate the emergent literacy skills which are considered to be the foundation of reading. The need for early and intensive intervention has been shown by the scientific literature indicating that the reading difficulties of many children can be prevented if early and intensive interventions are provided (Foorman & Torgesen, 2001).

A series of longitudinal studies have consistently found strong predictive relationships between early literacy skills and later performance in reading (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). The most important knowledge that is predictive of future reading ability has been well documented. To summarize, critical skills include the ability to identify and manipulate sounds in words (i.e. phonological and, more particularly, phonemic awareness) and the ability to associate printed letters with sounds in order to decode unfamiliar words (i.e. alphabetic principle). These literacy skills are particularly critical for low-SES children because they enter school at a disadvantage. More specifically, they tend to enter first grade performing below their peers from middle-income families (Clemens, Ragan, & Widales-

Benitez, 2016; Strang & Piasta, 2016).

Meta-analyses have shown that the systematic and explicit teaching of letter-sound correspondences and decoding procedures, together with the application of these skills in reading and writing activities, is the most effective method for improving literacy skills in children with reading disabilities (Ehri, Nunes, Stahl & Willows, 2001; McArthur, Eve, Jones, Banales, & Kohnen, 2012; Suggate, 2010; Galuschka, Ise, Krick, & Schulte-Körne, 2014). Moreover, authors argue that small-group reading interventions are more efficient than individual or whole class interventions (Ehri et al., 2001; Piasta & Wagner, 2010; Scammacca, Roberts, Vaughn, & Stuebing, 2015; Vaughn et al., 2010; Hall & Burns, 2018). However, this is inconsistent with the work of Vaughn et al. (2003) which shows that 1:1 and 1:3 interactions are the most effective. Interventions need to be sufficiently well targeted to address the specific reading deficits of the students in these small groups. Authors have shown that interventions are more effective if they are correctly targeted to address the specific reading deficits of the students in the group (Burns, VanDerHeyden, & Zaslofsky, 2014; Burns et al, 2016).

The use of targeted interventions requires a reliable and valid measure of the skills that contribute to reading acquisition. Many studies have focused on identifying and reliably measuring basic literacy skills in order to understand the reading process in typically achieving and struggling readers. In this study, we have used a research-based collection of individual subtasks that measure some of the foundational skills needed for reading acquisition in French. This instrument can serve as a baseline for early reading acquisition. It is similar to the *Early Grade Reading Assessment* (EGRA), which has been used in more than 70 countries for over 120 languages. (Dubeck & Gove, 2015; Stern, Dubeck, & Dick, 2018).

The purpose of our second study was to examine whether code-based teaching focusing on the explicit and systematic training of decoding, including letter recognition, letter-sound

correspondence, phonological awareness and fluency, impacts reading skills in first grade when pedagogical interventions are targeted toward children with difficulties.

Research questions

To summarize, two types of interventions could be effective in reducing children's difficulties during learning to read. The Ministry of Education can decide to drastically reduce class sizes. However, what is the impact of such policy measure on literacy skills? Alternatively, what is the impact of evidence-based educational interventions focusing on at-risk children administered by teachers in normally sized classes? Our aim is to present two large-scale studies in Grade 1 using the same statistical analyses as are adopted in observational studies.

Method

The observational studies were conducted in France and used the same conventional design: an experimental group in which the children were exposed to a specific intervention (policy or educational) and a control group for which no specific measure was applied. The children's literacy skills were assessed twice, i.e. before and after intervention. To examine the potential effect of such interventions, we calculated treatment-effects estimators that use matching (see below), since this kind of analysis is well suited to observational studies (Stuart, 2010).

Study 1: What are the effects when class size is reduced?

Participants

During the school year 2002-03, the French Ministry of Education decided to reduce class sizes specifically in areas exhibiting a variety of social difficulties, so-called *Zones d'Education Prioritaire* (ZEP; zones with specific educational needs). In a quasi-experiment, teachers and classrooms were first assigned to small and large class sizes. The children in normal classes (with 20-25 pupils, control group) were then compared with those in classes of small size (experimental group). These latter classes had no more than 12 pupils. One hundred

classes in each condition distributed throughout different regions of France participated in this research. The education authorities were committed to this experiment. New analyses are presented below on data collected from 1095 children (only those who did not repeat a year), i.e. 572 children in the experimental group and 523 in the control group. Their mean age was 76.5 months (sd = 3.5) at the beginning of the school year.

Literacy skills: material and procedure

The children's literacy skills were assessed by the teachers in two sessions. They first administered a pre-reading skills assessment at the beginning of Grade 1 (October-November; t1) and then a spelling and reading assessment at the end of the period (May-June; t2). These tasks, which were presented in booklets, were performed in small groups. However, the teachers did not record the results and external ratings were obtained.

Pre-reading skills (t1). Three short tasks were proposed to assess letter knowledge, phonological skills and high-frequency word recognition. A global score was calculated to estimate the children's performances in these literacy skills, which are well known to be closely related to later performance in reading and spelling.

Word spelling (t2). The children had to write 24 words underneath a picture named by the teacher. The words had a high frequency index and a spoken CV (e.g.: *main*, hand), CVCV (e.g.: *lapin*, rabbit) or CCVC (e.g.: *fleur*, flower) structure. For this task, we calculated a composite score which we refer to as the phonological recoding score. This score comprises both orthographically correct (e.g. *main*) and phonologically correct (e.g. *min*) responses.

Word reading (t2). In this forced-choice task, the same words as in spelling were proposed. The children had to find the target word in a list of four items consisting of the orthographically correct word (e.g., *lapin*), and three pseudowords, namely a homophone

lapain), a visually similar item (*lapiu*), and an item sharing the same initial letters (*lacet*). Here, we consider only the orthographically correct response.

Results

The descriptive data (Table 1) show scores on pre-reading skills at t1 in each group. These are very similar, whereas the spelling and reading scores at t2 were higher in the experimental group than in the control group. To examine this advantage, we calculated an estimator of treatment effects using nearest-neighbor matching (NNM). The general principle was to match individuals in the treatment group who were as close as possible to individuals in the no-treatment group before intervention. To do this, the scores at t1 were considered as covariates in order to select the nearest neighbors in each group. The average treatment effect (ATE) measures the difference in mean (average) outcomes between individuals assigned to the treatment and individuals assigned to the control group (see Abadie, Drukker, Herr, & Imbens, 2004)¹. We used Stata 14 for these analyses (Table 2). When ATE is positive, the effect is in favor of the experimental group.

[Insert Table 1]

[Insert Table 2]

¹ The nearest-neighbor matching is based on the Mahalanobis distance: $D_{ij} = (X_i - X_j)' \Sigma^{-1} (X_i - X_j)$. If the subject of interest is the ATT (average effect of the treatment on the treated), Σ is the variance-covariance matrix of X in the full control group; if it is the ATE, then Σ is the variance-covariance matrix of X in the pooled treatment and full control groups (Stuart, 2010, p. 6).

The results of NNM (Table 2) show a significant effect of class size in favor of the experimental group in reading, with an ATE of 0.54, as well as in spelling (1.41), with the effect size being higher in spelling.

Study 2: What are the effects of targeted educational interventions?

Participants

This large-scale study involved 2803 children *schooled in normal-size classes* who, as a quasi-experiment in Study 1, had been assigned to two groups: one experimental ($N = 489$ (248 boys; 241 girls) mean age = 75.5 months; $sd = 4.3$ m.), and one control ($N = 2314$ (1172 b; 1132 g; 10 not indicated) m. a. = 75.6 m.; $sd = 3.8$ m.). Only schools in ZEP (zones with specific educational needs) distributed across different regions of France participated in this study which was consented to by the academic authorities and which involved volunteer teachers. In the two groups, average size by class was very similar, namely 20 children in the experimental group and 20.1 children in the control group. Concerning the socio-economic status (SES), a balance was also observed between the two groups with regard to the number of children from low SES (Cont: 55%; Exp: 54%), and middle and high SES backgrounds (respectively 24% and 13% in the two groups).

Assessment: material and procedure

As in Study 1, the teachers contributed to the assessment of literacy skills by administering collective (Coll) tasks to the children in small groups during two sessions before (t1) and after (t2) the intervention. All the tasks were presented in booklets and a training item was proposed for each. The ratings were not performed by the teachers but by two independent raters. Moreover, two tasks at t2 were administered individually (Ind) to the children by different experimenters (see below). The two major components of reading were assessed using different tasks. These related 1/ to the code, involving letter knowledge, phonological

skills, word and pseudoword reading, word spelling, fluency and 2/ to comprehension, involving vocabulary, oral and written comprehension. Finally, performance on numeracy skills was also assessed. Indeed, if interventions on literacy skills are specifically aimed at children experiencing difficulties in this domain, we would not expect to observe any effect on numeracy skills.

Letter knowledge (t1; Coll). Two tasks were proposed: in the letter writing task, the children had to write the letter named by the teacher (max = 15). In the letter sound task, a pseudoword was first named by the teacher and the children then had to circle the first letter (among 7) of the pseudoword (max = 15).

Phonological skills (t1, t2; Coll). A phonemic segmentation task was administered: the children had to draw as many circles representing phonemes as they heard in the pseudoword named by the teacher (max = 12).

Word reading (t1, t2; Coll). A forced-choice task was proposed: the children had to circle the (orthographically correct) word corresponding to the picture and named by the teacher. For each item, five sequences of letters were proposed (words and pseudowords) (at t1: max = 10; at t2: max = 16).

Pseudoword reading (t1, t2; Coll). The same type of task (without picture) was administered (max = 10).

Word spelling (t2; Coll). Twelve words were dictated by the teacher. Only orthographically correct written words were taken into account (max = 12).

Vocabulary (t1; Coll). For each item, the children had to answer two questions: they had 1/ to circle the picture which, out of four, corresponded to the word spoken by the teacher and 2/ to answer a question (yes or no) about the word by circling the smiling face for a "yes" response or the frowning face for a "no" response (e.g.: 1/ *you have to circle the raft*; 2/ *does a raft walk on paths?*) (max = 26).

Listening comprehension (t1; Coll). The children listened twice to a short narrative (110 words) and then had to answer questions by circling the picture (out four) which they thought was correct.

Reading comprehension (t2; Coll). Short sentences were presented above four pictures. The children had to circle the picture corresponding to the sentence (max = 9).

Reading comprehension (t2; Ind). The children were asked to read a very short text (47 words) aloud. They were then asked to read questions about the text and to answer the experimenter orally (max = 5).

Fluency (t2; Ind). The children had to read a short text (50 words). The experimenter recorded the time taken to read the text as well as the number of errors. Fluency was calculated using the following formula: $(50 - \text{Errors}) / (\text{Time in seconds} * 60)$.

Numeracy skills (t1). A series of numbers or geometrical figures was presented in a forced-choice task and the children had to circle the item that had been named by the teacher (max=10).

Numeracy skills (t2). Four tasks were administered. The first consisted in writing a number (/5). In the second task, the children had to write the number missing from a series of numbers (/6). For the third task, the teacher orally presented a short problem in an additive or subtractive situation and the children had to give their response by circling the correct number in a series of proposed numbers (/5). In the fourth task, the children had to complete additive operations involving numbers lower than 30 (/4). We considered the total score (max = 20).

Educational interventions

A booklet presenting a program centered on code learning was presented at a one-day training course to the teachers who had enrolled in the experimental group. The program, together with all the linguistic material and detailed exercises, included four main levels: training

phonological skills (1), learning letter sounds (2) decoding (3) and fluency (4). For the first level, various exercises were proposed: word segmentation into syllables, deleting syllables, segmentation of short pseudowords into phonemes, blending phonemes, etc. Once the children had reached a high level of performance in terms of phonological skills, they learned letter-sound correspondences (2nd level) involving the ten high-frequency letters. When they had acquired a sufficient level of skill with the learned letters, they were trained (3rd level) in decoding CV, CVC, CVCV pseudowords. Finally, fluency (4th level) was trained through the repeated reading of short texts during which the children were encouraged to read as quickly and accurately as possible. The teachers were recommended to dedicate 30 min a day in small groups (between 10-12 children) for the children with average or high levels of literacy skills, whereas for the lowest performing children one hour (2*30 min a day) was advised (4-6 children per group). The teachers were supported by educational aids only during group teaching. This protocol was constructed by the Association Agir pour l'Ecole (APE; Act for School), a platform that promotes experimentation in the use of new evidence-based methods of learning to read. Compliance with the protocol was monitored by experimenters from APE, who performed weekly follow-up.

Before the training, the children were assessed on certain classical pre-reading skills (phonology, letter knowledge, etc.). Depending on their initial level, they were assigned to a group ranging from Group 1 (good level) to Group 5 (poor level). The amount of training they received therefore depended on their initial level: the lower this level was, the more training they received².

² A paper on the impact of the period of training on later reading and spelling performance is currently being prepared.

Results

In Table 3, we observe that the differences between mean scores in the two groups vary at t1, sometimes being in favor of the experimental group (letter sound, phonemic segmentation, pseudoword reading), sometimes in the opposite direction (letter writing), and sometimes being very small (word reading, vocabulary, listening comprehension, numeracy). However, the differences at t2 are systematically in favor of the experimental group (not for numeracy). To control for these differences in initial literacy skills between groups at t1, we again used the nearest-neighbor matching analysis with specific covariates at t1 according to the dependent variable analyzed at t2. The insertion of covariates in the NNM depends on the type of task, which may involve the processing of short units (i.e. phonemic segmentation), the reading of words and pseudowords, or fluency, reading comprehension or numeracy. Results are presented in Table 4.

[*Insert Table 3*]

[*Insert Table 4*]

All ATE on literacy skills are positive, meaning that the intervention was beneficial to the experimental group in all the tasks related to literacy. However, significant effects are observed only in word reading, in word spelling, in reading comprehension (in the two tasks, administered collectively and individually), and in fluency. The effect sizes vary from .12 (word spelling) to .32 (in reading comprehension administered individually). Only one ATE is negative, namely on numeracy skills. However, this effect is far from being significant, meaning that the small difference observed is negligible.

Discussion

In Study 1, we also found, as reported in the literature, that the students in reduced classes performed better and exhibited significant effects in terms of literacy skills when compared to those who attended full-size classes. This result was obtained by means of a specific statistical analysis, namely a matching technique that provides treatment-effects estimators and is now acknowledged to be highly suitable for observational studies (Stuart, 2010). The effect sizes observed in word reading and word spelling are, however, relatively small and are very similar (or the same: .22) as reported by Bressoux (2016). Clearly, the policy intervention, which aimed at reducing the size of the classes, did have an effect, albeit rather small (as expected in large-scale studies; see below), on performance in two major literacy skills, i.e. reading and spelling at the end of first grade. Moreover, this small effect must be considered in the light of the high cost to the ministry of increasing the number of first grade teachers in disadvantaged areas. In other words, the cost/effectiveness ratio is low compared to specific educational interventions in full-size classes.

In Study 2, which was again conducted with children in disadvantaged areas, the educational interventions had two characteristics: They were focused uniquely on the code and were targeted at children at risk of reading failure as suggested by their initial literacy skills. The lower this initial level was, the more time the teachers spent trying to reduce the children's difficulties in small groups. Globally, the results show significant effects on several literacy skills: word reading, word spelling, reading comprehension, and fluency. The treatment-effects estimators show that the performance of the students in the experimental classes improved more than that of those in the control classes. We can again note that the impact of these interventions is small, with effect sizes ranging from low for word reading (.12) to moderate (.32) for comprehension (in one task). Unexpectedly, these targeted interventions also had a significant effect on reading comprehension despite the fact that meaning-based

interventions were not specifically implemented during these pedagogical interventions. However, within the perspective of the Simple View of Reading, we could expect code-based interventions to impact word-reading and automate this process, thereby freeing up cognitive resources that can then be allocated to comprehension.

Finally, the absence of a significant effect in mathematics constitutes an argument in favor of targeted code-focused teaching. Indeed, if a general teacher effect were to have a global impact on teaching then we should also have observed an effect on performance in mathematics. Clearly, this was not the case.

Why did we observe such effect sizes in the two studies? Slavin and Smith (2009) found a negative relationship between effect sizes and sample sizes: The bigger the samples, the smaller the effect sizes. More recently, in a more complete paper, Cheung and Slavin (2016) analyzed the methodological features that could affect effect size in educational studies on training. Two factors are of interest for us: sample size and researcher-made measures vs independent measures³. When the authors compared effect sizes found in small samples (≤ 250) vs large samples (> 250), they found that the average effect size was almost twice as large (+.30) in small samples than in large samples (+.16). Moreover, when they compared the average effect size in the smaller samples (< 100) with that in the largest ones (> 2000), they found that the two differed by a factor of more than 3.5 (+.38 vs +.11). Moreover, when they compared the kind of measures used in the studies, they found that the effect size was twice as great (+.40) in researcher-made measures than in independent measures (+.20). In the

³ The other factors which significantly impact effect sizes are also presented in their paper: a/ randomized (+0.16) vs quasi-experiments (+0.23), b/ elementary (+0.20) vs secondary studies (+0.17), at limit of significance, and c/ publication bias with effect size higher in published articles (+.30) than in unpublished reports (+.16).

light of these two meaningful results, we have to examine our results in more detail. Indeed, we found effect sizes greater than the average effect size calculated in large samples by Cheung and Slavin (2016). In Study 1, our effect sizes (.14 and .22) were greater than the average effect size found by the above authors in samples from 1000 to 1999 (+.13) and all our effect sizes in Study 2 (from .12 to .32) were greater than the average effect size (+.11) found by the same authors in samples ≥ 2000 . However, our effect sizes should be considered in relation to the second factor that impacts effect sizes: They were found with our own measures, i.e. researcher-made measures, namely the condition in which Cheung and Slavin reported larger effect sizes. We should perhaps consider the possibility that our results are similar to the standard values reported in the literature.

However, one of the effect sizes, .32 in reading comprehension (Study 2) could be considered rather large. Even though no specific training in comprehension was provided, this gain could be explained by the training in fluency. Indeed, on the one hand, correlational studies have revealed developmental links between fluency and comprehension in Grades 1 and 2 (Kim, Wagner, & Lopez, 2012). On the other, to examine the potential causal link, Calet, Gutiérrez-Palma and Defior (2017) recently trained two groups of children (in Grades 2 and 4) in fluency and compared these with a control group (G1) without any specific training. One of the trained groups was trained with fluency considered as automaticity (accuracy and speed; G2) and the other was trained in automaticity and prosody (G3). The authors observed positive and significant performance improvements in reading comprehension in G2 and G3, with the improvement being greatest in G3. In our Study 2, only automaticity was trained and future research will need to take account of prosody during the training of fluency.

In research such as this which focuses on students' literacy performance, we have no information about other factors which might impact the performance gains resulting from the intervention. [It is one of the limitations of this research because the matching procedure is](#)

based exclusively on reading outcomes without including important covariates (e.g., teacher practice, parent and family background) that might affect the effects of class size and reading intervention". In fact, three categories of factors could be put forward: factors relating to children's engagement and attention, factors relating to teacher-pupil interactions, and factors relating to the teachers' experience. With regard to the first set of factors, it has been observed that pupils' inattentiveness and off-task behavior is reduced in small classes (Blatchford, Edmonds, & Martin, 2003) and that their engagement during learning is higher (Bressoux & Lima, 2004). It also seems likely that in small groups in experimental classes as in Study 2, pupils are more attentive during the targeted actions undertaken by teachers. Another limitation of our two studies is the absence of external observations during the lessons focused on code learning. In a reduced class-size or a class with small groups, we would expect the teaching practices and the nature of the teachers' interventions with the pupils to be changed by the context. Clanet (2010) observed that teachers acted differently when introducing new texts to read in small classes compared to normal classes. Moreover, in our Study 1, we had no information about teacher-student interactions or the time spent on interaction with individual children in small classes and large classes. Such information would have been extremely useful. The third category of factors, namely those associated with the children and the teachers, could not be analyzed (and, in particular, introduced as covariates) because we were not able to obtain the corresponding data: Family SES in the case of the children, and, with regard to the teachers, experience of teaching, age, initial and ongoing training, commitment and job satisfaction could also explain a part of the variance in the children's performance gains. Finally, the inference generated in this research paper may apply more specifically to a population of low-SES children given the specific sample characteristics of our two studies".

Researchers have conducted numerous studies into the effects of class size on children's reading achievement. However, the disparity in their results makes it difficult to draw overall conclusions. Few studies have focused on the class size effect on reading skills in first grade. Some studies have shown that class size makes a difference in early reading performance. Achilles (1996) suggests that the Project STAR data provides evidence that smaller classes lead to improved reading achievement. A large-scale study (NICHD, 2004) evaluated the extent to which first-grade class size predicted children's outcomes and observed classroom processes for 651 children. The results showed that first graders in smaller classes performed better on literacy skills. Moreover, smaller classrooms were found to provide higher quality instructional and emotional support.

Studies comparing class sizes of approximately 15 with those of around 25 in the early elementary grades reveal that class size has a significant impact on reading achievement, especially when teachers also make use of more effective instructional strategies. Mueller (2013) analyzed teacher experience as a moderating factor for the effect of class size reduction on student achievement. The main finding is that in the early grades, only experienced teachers are able to generate a beneficial class size effect on average student achievement. Moreover, smaller classes may allow teachers to give more individualized attention to their pupils and utilize very small group instruction more effectively. Thus, class size can make an even bigger difference when teachers change their teaching methods to suit smaller groups.

Globally, our findings in Study 2 are consistent with previous research regarding the importance of targeted reading interventions and with the previous studies of small-group interventions that have reported small effects (Vaughn et al., 2010). This study conducted with French-speaking children confirms that when children at risk of reading difficulties are identified early and receive appropriate interventions, many pupils acquire the skills necessary

to become successful readers. Although there is no single "right way" to teach pupils who have reading difficulties in first grade, researchers have identified key characteristics related to improved outcomes (Foorman & Torgesen, 2001; National Reading Panel, 2000; Torgesen, 2004; Wanzek & Vaughn, 2007). Interventions targeting a specific reading skill area have been found to be more effective than more general interventions that combine multiple reading skills. It is therefore vital to assess pupils in order to identify their needs and provide them with explicit instruction in the corresponding areas. Evidence suggests that increasing the intensity of effective instruction (use of smaller groups and more time spent on interventions) may have positive effects on student outcomes. In a recent meta-analysis, Hall and Burns (2018) suggest that targeted interventions and group size are the two most important intervention variables among those examined. One study (Vaughn et al., 2003) systematically studied the size of reading intervention groups. The effects of three grouping formats - 1:1 (one teacher with 1 pupil), 1:3 (one teacher with 3 pupils), and 1:10 (one teacher with 10 pupils) - on the reading outcomes of struggling second-grade readers were investigated. The children made significant gains in phoneme segmentation, fluency, and comprehension following the intervention, and these persisted 4-5 weeks after intervention. Based on effect sizes, both 1:1 and 1:3 were highly effective intervention group sizes for targeted reading instruction. In summary, providing targeted reading interventions to students in small groups appears to be an effective method for increasing pupils' reading skills. Wanzek and Vaughn (2007) summarized reading-intervention studies published between 1995 and 2005 that analyzed the effects of extensive interventions on reading outcomes for children in kindergarten through third grade. The results indicated that larger effects were reported by studies in which intervention was provided one-to-one or in small groups than by those in which intervention was provided in larger groups. We do not yet know with confidence the largest group size associated with positive effects.

Conclusion

To summarize, these two large-scale studies confirm the findings which have been reported in the literature. Although the effects of class size on literacy skills appear small, this must be considered as being due to the large samples used in the two studies. Moreover, these effect sizes must be assessed in the light of the associated financial cost. In Study 2, the effects of targeted pedagogical interventions could be considered as important and the economic cost is very low. Indeed, it is also necessary to consider the cost-effectiveness of such interventions. It is clear that policy interventions have a high economic cost (two teachers for 20-25 pupils in small classes against only one in full-size classes), while the cost of educational interventions is very low: We have to consider that teachers require a training period in order to adapt their educational interventions for at-risk children. Indeed, they have to be trained to assess reading ability and its related skills accurately, and to use the results of this assessment to implement practices that target students' specific needs in the light of their initial level. Finally, a double and simultaneous intervention, i.e. at both policy and educational level, could bring about a greater increase in the performance of students in disadvantaged areas. A drastic reduction in class size coupled with targeted interventions conducted in small groups in the light of the children's reading difficulties would probably be one of the best ways to cut the proportion of poor readers. Moreover, the use of digital devices, such as tablets, during learning to read could be a good way to stimulate the acquisition of this skill. New studies designed to investigate this new teaching context (with and without digital tools) are currently being conducted in France.

References

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, 4(3), 290-311.

- Achilles, C. M. (1996). Students Achieve More in Smaller Classes. *Educational Leadership*, 53(5), 76-77.
- Achilles, C. M., Harman, P., & Egelson, P. (1995). Using research results on class size to improve pupil achievement outcomes. *Research in Schools*, 2(2), 23-30.
- Anderson, L. W. (2000). Why should reduced class size lead to increased student achievement? In M. C. Wang, & J. D. Finn (Eds.), *How small classes help teachers do their best* (pp. 3-24). Temple University Center for Research in Human Development and Education: Philadelphia, PA.
- Blatchford, P., & Russell, A. (in press a). New ways of thinking about research on class size: an international perspective. *International Journal of Educational Research*, doi.org/10.1016/j.ijer.2018.09.011
- Blatchford, P., & Russell, A. (in press b). Class size, grouping practices and classroom management, *International Journal of Educational Research*, doi.org/10.1016/j.ijer.2018.09.004
- Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction*, 21(6), 715-730. doi.org/10.1016/j.learninstruc.2011.04.001
- Blatchford, P., Edmonds, S., & Martin, C. (2003). Class size, pupil attentiveness and peer relations. *British Journal of Educational Psychology*, 73, 15-36. Doi: 10.1348/000709903762869897
- Blatchford, P., Goldstein, H., Martin, C., & Browne, W. (2002). A study of class size effects in English school reception year classes. *British Educational Research Journal*, (28-2), 169-185. Doi: 10.1080/0141192012012213 0

- Blatchford, P., Moriarty, V., Edmonds, S., & Martin, C. (2002). Relationships between class size and teaching: A multi-method analysis of English infant schools. *American Educational Research Journal*, 39, 101-132. Doi: 10.3102/00028312039001101
- Bressoux, P. (2016). Research on class size in France. In P. Blatchford, K. Wing Chan, M. Galton, K. Chan Lai, & J. Chi-Kin Lee (Eds.), *Class size. Eastern and western perspectives* (pp. 80-91), London: Routledge.
- Bressoux, P., & Lima, L. (2004). *Evaluation de l'expérimentation CP à effectifs réduits. Détermination de l'implication des élèves et effets de l'implication des acquis*. Rapport pour le Ministère de l'Education Nationale. Grenoble: Université Mendès-France.
- Burns, M. K., Maki, E. E., Karich, A. C., Hall, M., McComas, J. J., & Helman, L. (2016). Problem-analysis at tier 2: Using data to find the category of the problem. In S. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 293-308). (2nd ed.). New York: Springer.
- Burns, M. K., VanDerHeyden, A. M., & Zaslofsky, A. F. (2014). Best practices in delivering intensive academic interventions with a skill-by-treatment interaction. In P. L. Harrison, & A. Thomas (Eds.). *Best practices in school psychology: Student-level services* (pp. 129-141). Bethesda, MD: National Association of School Psychologists.
- Calet, N., Gutierrez-Palma, N., & Defior, S. (2017) Effects of fluency training on reading competence in primary school children: The role of prosody. *Learning and Instruction*, 52, 59-68. doi.org/10.1016/j.learninstruc.2017.04.006
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283-292. Doi: 10.3102/0013189X16656615

- Clanet, J. (2010). The relationship between teaching practices and student achievement in first year classes: A comparative study of small size and standard size classes. *European Journal of psychology of Education, 25*, 192-206. Doi: 10.1007/s10212-010-0012-y
- Clemens, N. H., Ragan, K., & Widales-Benitez, O. (2016). Reading difficulties in young children. *Behavioral and Brain Sciences, 3*(2), 177-184. Doi: 10.1177/2372732216656640.
- Ding, W., & Lehrer, S.F. (2010). Estimating treatment effects from contaminated multi-period education experiments: The dynamic impacts of class size reductions. *The Review of Economics and Statistics, 92*(1), 31-42. doi.org/10.1162/rest.2009.11453
- Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development, 40*, 315-322. <http://dx.doi.org/10.1016/j.>
- Ecalte, J., Magnan, A., & Gibert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. *Journal of School Psychology, 44*, 191-209. doi:10.1016/j.jsp.2006.03.002
- Ehrhardt J., Huntington N., Molino J., Barbaresi, W. (2013) Special education and later academic achievement. *Journal of Developmental and Behavioral Pediatrics, 34*, 111-119. Doi: 10.1097/DBP.0b013e31827df53f.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., Willows, D., M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*, 393-447. Doi: 10.3102/00346543071003393
- Finn, J. D., & Achilles, C. M. (1999). Tennessee's Class Size Study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*, 97-109. Doi: 10.3102/01623737021002097

- Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97*(2), 214-223. doi.org/10.1037/0022-0663.97.2.214
- Foorman, B. R., & Torgesen, J. K. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice, 16*(4), 202-211. Doi: 10.1111/0938-8982.00020
- Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*(2-3), 613-639.
- Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37-55.
- Galuschka, K., Ise, E., Krick, K., Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE, 9*(2), e89900. doi:10.1371/journal.pone.0089900.
- Goldstein, H., & Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal, (24-3)*, 255-268.
- Hall, M. S., & Burns, M. K. (2018). Meta-analysis of targeted small-group reading interventions. *Journal of School Psychology, 66*, 54-66. doi.org/10.1016/j.jsp.2017.11.002
- Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research, 43*(6), 387-425. doi.org/10.1016/j.ijer.2006.07.002

- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
- Hattie, J. (2012). *Visible learning for teachers*. London, UK: Routledge.
- Kim, Y. S., Wagner, R. K., Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology*, 113, 93-111. <http://dx.doi.org/10.1016/j.jecp.2012.03.002>
- Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *The Elementary School Journal*, 108(4), 275-291. doi.org/10.1086/528972
- Konstantopoulos, S. (2011). How consistent are class size effects? *Evaluation Review*, 35(1), 71-92. Doi: 10.1177/0193841X11399847
- Konstantopoulos, S., & Sun, M. (2014). Are teacher effects larger in small classes? *School Effectiveness and School Improvement*, 25(3), 312-328.
- Konstantopoulos, S., & Traynor, A. (2014). Class size effects on reading achievement using PIRLS data: Evidence from Greece. *Teachers College Record*, 116(2), 1-29.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114, 497-532. doi.org/10.1162/003355399556052
- McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., et al. (2012). Phonics training for English-speaking poor readers. *The Cochrane Database of Systematic Reviews*, 12: CD009115. Doi: 10.1002/14651858.CD009115
- Mueller, S. (2013). Teacher experience and the class size effect. Experimental evidence *Journal of Public Economics*, 98, 44-52. doi.org/10.1016/j.jpubeco.2012.12.001

- NICHD (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology*, 40(5), 651-664. Doi: 10.1037/0012-1649.40.5.651
- NRP National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, D.C.: National Institute of Child Health and Human Development.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2), 127-142. Doi: 10.3102/01623737021002127
- Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, 37, 123-151. Doi: 10.3102/00028312037001123
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2002). Do low-achieving students benefit more from small classes? Evidence from the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 24(3), 201-217. Doi: 10.3102/01623737024003201
- Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, 45, 8-38. doi.org/10.1044/1092-4388(2010/09-0093
- Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12: 1980–2011. *Journal of Learning Disabilities*, 48, 369-390. Doi: 10.1177/0022219413504995

- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265-282.
- Shen, T., & Konstantopoulos, S. (2017). Class size effects on reading achievement in Europe: evidence from PIRLS. *Studies in Educational Evaluation, 53*, 98-114. doi: 10.1016/j.stueduc.2017.04.001.
- Slavin, R. E., Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic review in education. *Educational Evaluation and Policy Analysis, 31*(4), 500-506. Doi: 10.3102/0162373709352369
- Stern, J. M. B., Dubeck, M. M., & Dick, A. (2018). Using Early Grade Reading Assessment (EGRA) data for targeted instructional support: Learning profiles and instructional needs in Indonesia. *International Journal of Educational Development, 61*, 64-71. doi.org/10.1016/j.ijedudev.2017.12.003
- Strang, T., & Piasta, S. B. (2016). Socioeconomic differences in code-focused emergent literacy skills. *Reading & Writing, 29*, 1337-1362.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-21. doi: 10.1214/09-STS313.
- Suggate, S. P., (2010). Why what we teach depends on when: grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*, 1556-1579.
- Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7-26. doi.org/10.1016/S0022-4405(01)00092-9
- Torgesen, J. K. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. *The voice of evidence in reading research, 355-382*.

- Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*(5), 301-315. Doi: 10.1177/07419325030240050501
- Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. D., et al. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39*, 3-21.
- Vellutino, F. R., & Zhang, H. (2008). Preventing long-term reading difficulties through kindergarten and first grade intervention: The case for early intervention. *Perspectives on Language Learning and Education, 15*(1), 22-33. Doi:10.1044/lle15.1.22
- Wanzek J, Vaughn S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541-561.
- Zyngier, D. (2014). Class size and academic results, with a focus on children from culturally, linguistically and economically disenfranchised communities, *Evidence Base, 1*, *journal.anzsog.edu.au, ISSN 1838-9422*.

Table 1: Mean scores (standard deviations) on literacy skills (with their reliability coefficients) at t1 and t2 in Experimental and Control groups.

| | Evaluation time | Exp group (N=572) | Cont group (N=523) | Reliability (N=1095) |
|-------------------|-----------------|----------------------|-----------------------|-------------------------|
| Pre-reading (/28) | t1 | 16.69 (5.10) | 16.44 (6.53) | .61 ^a |
| Reading (/24) | t2 | 15.4 (4.87) | 14.72 (5.58) | .81 ^b |
| Spelling (/24) | t2 | 15.34 (7.22) | 13.72 (7.58) | .91 ^b |

Notes. ^a: split-half reliability coefficient; ^b: Cronbach alpha

Table 2: Results of nearest-neighbor matching

| DV at t2 | Covariate at t1 | ATE CI 95% | Std Error | z p | effect size |
|----------|-----------------|---------------|-----------|--------|-------------|
| Reading | | 0.54 | 0.27 | 1.96 | .14 |
| | pre-reading | 0.00 – 1.07 | | .05 | |
| Spelling | | 1.41 | 0.37 | 3.78 | .22 |
| | | 0.68 – 2.14 | | .0001 | |

Notes. ATE: average treatment effect; formula of effect size: $(m^t - m^c) / \sigma^t$ (Stuart, 2010) where m represents the means in the treatment and control group.

Table 3: Mean scores (standard deviations) in literacy and numeracy skills (with their reliability coefficients) at t1 and t2

| | Evaluation time | Exp group (N=489) | Cont group (N=2314) | Reliability ^a (N=2803) |
|----------------------------------|--------------------|----------------------|------------------------|--------------------------------------|
| Letter writing (/15) | t1 | 12.03 (3.78) | 12.41 (3.60) | .91 |
| Letter sound (/15) | t1 | 12.56 (3.69) | 12.18 (3.83) | .92 |
| Phonemic segmentation (/12) | t1 | 5.46 (2.92) | 4.62 (3.05) | .82 |
| | t2 | 8.33 (3.45) | 7.73 (3.67) | .89 |
| Word reading | t1 (/10) | 2.67 (1.91) | 2.76 (2.06) | .61 |
| | t2 (/16) | 13.35 (3.72) | 12.80 (4.29) | .93 |
| Pseudoword reading (/10) | t1 | 6.28 (2.49) | 5.93 (2.62) | .80 |
| | t2 | 8.50 (2.40) | 8.23 (2.62) | .88 |
| Word spelling (/12) | t2 | 8.93 (3.06) | 8.50 (3.48) | .88 |
| Vocabulary (/26) | t1 | 19.90 (4.39) | 19.74 (4.80) | .88 |
| Listening comprehension (/12) | t1 | 8.13 (2.53) | 8.19 (2.61) | .81 |
| Reading comprehension C (/9) | t2 | 6.28 (2.29) | 5.94 (2.54) | .82 |
| Reading comprehension I (/5) | t2 | 3.44 (1.56) | 2.87 (1.76) | .88 |
| Fluency | t2 | 59.00 (34.55) | 52.32 (34.64) | – |
| Numeracy skills | t1 (/10) | 7.52 (2.50) | 7.72 (2.44) | .82 |
| | t2 (/20) | 14.47 (4.90) | 14.68 (4.91) | .91 |

Notes. C: collectively; I: individually; ^a: Cronbach's alpha.

Table 4: Results of nearest-neighbor matching

| DV at t2 | Covariates at t1 | ATE (CI 95%) | Sd Error | z <i>p</i> | Effect size |
|-------------------------|---------------------------------------|----------------------|----------|---------------|-------------|
| Phonemic segmentation | Letter writing | .29 | .26 | 1.21 | – |
| | Letter sound Phonemic segmentation | (-.18 – .23) | | ns | |
| Word reading | Letter writing | .53 | .19 | 2.79 | .13 |
| | Letter sound | (.16 – .90) | | .005 | |
| | Phonemic segmentation | | | | |
| | Word reading | | | | |
| Pseudoword reading | Letter writing | .24 | .15 | 1.63 | – |
| | Letter sound | (-.05 – .53) | | ns | |
| | Phonemic segmentation | | | | |
| | Pseudoword reading | | | | |
| Word spelling | Letter writing | .48 | .17 | 2.78 | .12 |
| | Letter sound | (.14 – .82) | | .005 | |
| | Phonemic segmentation | | | | |
| Reading comprehension C | Vocabulary | .44 | .12 | 3.82 | .14 |
| | Listening comprehension | (.21 – .67) | | .0001 | |
| Reading comprehension I | Vocabulary | .74 | .08 | 8.81 | .32 |
| | Listening comprehension | (.57 – .90) | | .0001 | |
| Fluency | Letter writing | 8.10 | 1.83 | 4.43 | .19 |
| | Letter sound | (4.51 – 11.67) | | .0001 | |
| | Phonemic segmentation | | | | |
| | Word reading | | | | |
| Numeracy skills | Numeracy skills | -.10 (-.52 – .32) | .21 | -0.49 ns | – |

Notes. ATE: average treatment effect; C: collectively; I: individually.