June 2022

The Combined Effects of Regular Classroom and Remedial Learning Programs on the Reading Achievement of First Grade Pupils

STEVE BISSONNETTE
Professeur titulaire au département d'éducation de la TÉLUQ

CHRISTIAN BOYER
Consultant en pédagogie et en orthopédagogie, SESSIONS

FRÉDÉRIC MORNEAU-GUÉRIN
Professeur adjoint au département d'éducation de la TÉLUQ

Collection
Des points sur les i et des barres sur les t

Éditions de l'Apprentissage
www.editionsdelapprentissage.com

For You, My Love

I went to the market of birds
And I bought birds
For you
My love

I went to the market of flowers
And I bought flowers
For you
My love

I went to the market of ironwork
And I bought chains
Heavy chains
For you
My love

And then I went to the market of slaves
And I looked for you
But I did not find you there
My love

Jacques Prévert, poem published in
1946 in the collection *Paroles*

Pour toi mon amour

Je suis allé au marché aux oiseaux
Et j'ai acheté des oiseaux
Pour toi
Mon amour

Je suis allé au marché aux fleurs
Et j'ai acheté des fleurs
Pour toi
Mon amour

Je suis allé au marché à la ferraille
Et j'ai acheté des chaînes
De lourdes chaînes
Pour toi
Mon amour

Et puis, je suis allé au marché aux esclaves
Et je t'ai cherchée
Mais je ne t'ai pas trouvée
Mon amour

Jacques Prévert, poème publié en 1946
dans le recueil *Paroles*

 ISBN : 978-2-923805-66-5

Bissonnette, S., Boyer, C., Morneau-Guérin, F. (2022). The Combined Effects of Regular Classroom and Remedial Learning Programs on the Reading Achievement of First Grade Pupils. Montréal : Éditions de l'Apprentissage.

Abstract

In 2003, two school-based programs for teaching reading in English were introduced in Québec: the Accelerated Development of Reading (ADOR) program for first graders and, for ADOR pupils with weak reading skills, the Intensive Intervention in Reading (IIR) support program. Explicit and direct instruction provides the framework for ADOR and IIR. Two quasi-experimental studies were conducted to measure the effects of the ADOR and IIR programs compared to regular reading instruction and remediation. In both studies, the combined effects of ADOR and IIR were measured on two separate cohorts of grade 1 pupils over a full school year. This study shows that the ADOR and IIR programs result in superior reading performance and the largest effects were on subtests measuring reading comprehension.

Keywords: first grade, reading program, reading comprehension, explicit instruction, direct instruction.

www.editionsdelapprentissage.com
Dépôt légal : 2e trimestre 2022
Bibliothèque et Archives nationales du Québec
ISBN : 978-2-923805-66-5
Version 10/06/22

 ISBN : 978-2-923805-66-5

Introduction

In the spring of 2003, two school-based programs for teaching reading in English were introduced in Québec: the *Accelerated Development of Reading* (ADOR) program for first graders and, for ADOR pupils with weak reading skills (as well as students in higher grades), the *Intensive Intervention in Reading* (IIR) support program. These two programs were translated and adapted from French-language programs in use in Québec: *Développement accéléré de la lecture en première année* (DAL) and *Développement intensif du raisonnement* (DIR) in reading (known in its early days as *Intervention intensive en lecture*). It should be noted that the translation and adaptation into English of educational programs and materials, originally written and designed in French, is a rather unusual phenomenon. French versions of these programs began to take shape in Estrie in the fall of 1990.

The design of these programs draws largely on evidence from the past and is supported by present evidence. Explicit and direct instruction (Boyer, 1993) provides the framework for ADOR and IIR while incorporating the other components that will be described below. Moreover, their development and dissemination perfectly align with both the spirit and the letter of the Rational Results-Based Management (RRBM) method, as described in Boyer and Bissonnette (2021).

In 2010, Boyer published a book presenting the results of the DIR program in reading (Boyer, 2012). Dion (2012) and Forget (2012) refer to this book, pointing out that although the DIR program is evidence-based, there has been—to their knowledge—no experience-based validation of the program. It is important to note that several years earlier, two quasi-experimental studies were conducted to measure the effects of the ADOR and IIR programs (Savage, 2005; Savage & Deault, 2007). They were conducted by an external team of researchers on behalf of an English school board that was experimenting with ADOR and IIR. It is our understanding that, regrettably, the results from Savage (2005), and from Savage and Deault (2007) have been neither disseminated nor resulted in any official publication, other than the research team's report to the school board. In both studies, the combined effects of ADOR and IIR were measured on two separate cohorts of grade 1 pupils. The first and second measurements, in 2004—2005 and 2006—2007 respectively, implemented the programs over a full school year. However, in the coming pages, we will present only the results from the 2007 report (Savage & Deault, 2007), as it includes the results from the 2005 report[1].

Methodology

As noted earlier, external researchers who were not the program designers[2] conducted two studies, the results of which we present below. Students' reading skills in English were assessed at the beginning (pretest) and at the end (posttest) of the school year, using the *Group Reading Assessment and Diagnostic Evaluation* (GRADE)[3] as a standardized measure. An arithmetic measure, the *Wide Range Achievement Test* (WRAT), was also included to control for the Hawthorne effect[4].

The experiment took place over two full school years (2004-2005 and 2006-2007). In 2004-2005, the explicit ADOR-IIR instructional programs described below were tested with 117 children from 6 classes in 2 schools. In 2006-2007, the ADOR-IIR programs were tested again with 87 pupils from 4 classes in the same 2 schools as the previous year. Two teachers participated in both experiments. A total of 204 pupils experienced the ADOR-IIR programs and made up the experimental group. The control group consisted of 276 children from 13 classes in 7 schools who received the standard reading instruction described below. Reading instruction in both the experimental and control groups was provided by teachers who had between 1 and 37 years of teaching experience. There were no significant differences in the average number of years of teaching experience between teachers in the experimental and control groups. School's district administrators selected which schools would take part, and teachers agreed to participate in the study after an information session.

 ISBN : 978-2-923805-66-5

The ADOR-IIR experimental group: Implementing the ADOR program in regular classes with all first-grade pupils

Intended for all first-grade students, the ADOR program aims to develop children's reading skills in their first 80 to 100 days of school. Both the ADOR program (the English-language version) and the DAL program (the French-language version) draw on evidence of effective instruction in learning to read (National Reading Panel, 2000), Explicit Teaching of reading comprehension (Boyer, 1993), and Direct Instruction (Carnine, Silbert & Kameenui, 1997), as well as on research on resilience and optimism (Seligman, 1995), Kirschner's (2002) work on cognitive load, and research on behavioural interventions and token-saving systems that promote effort and academic achievement (Little & Akin-Little, 2003). The ADOR, like its French-language counterpart, is clearly an explicit and direct instructional program that emphasizes the development of reasoning while reading, skills for retrieving information in a read text, systematic acquisition of decoding, and the increase in fluency, including prosody.

ADOR is also consistent with current evidence recommending the development of vocabulary and general world knowledge, which are recognized as crucial to reading comprehension (Oakhill, Cain, & Elbro, 2015; Stuart & Stainthorp, 2016; Willingham, 2017).

At the beginning of the learning process, the ADOR program includes doing daily phonological-skill activities, learning of phonemes systematically, decoding simple words, understanding sentences from short texts read to or with pupils, reasoning while reading texts or sentences read to or with pupils, and developing vocabulary. After 14 days of instruction, some of the most common letter combinations in English are systematically introduced (e.g., ay, ea, oo and sh), letter activities are greatly reduced, and phonological activities and some decoding activities are omitted altogether. Reasoning, comprehension, and vocabulary activities are retained throughout the year. An oral reading fluency activity is incorporated into daily activities when pupils[5] can read aloud at an exact-rate of 12 to 19 words per minute.

Brief and frequent formative assessments are provided throughout the program, among other objectives, to identify the weakest pupils and to adjust instruction and activities to the group's average performance. The two to five pupils in the class with the weakest reading skills then receive additional ad hoc support from the teacher, called *ad hoc* remediation. This help is provided individually or in small groups for 5-10 minutes per day. After approximately 10 days, the pupils are re-assessed. The results of the second evaluation determine whether the intervention was effective and whether the student will continue to receive *ad hoc* support from the teacher.

At the beginning of the school year, ADOR classroom teachers attended in a four-day training session given by the school board's trained coaches. After the training session, the coaches demonstrated specific pedagogical activities and classroom management techniques in regular classes (with the pupils and their teacher) and observed the extent to which the teachers were applying the content from the training sessions and the coach's demonstrations. Observation grids are used and teachers received feedback. Each teacher was observed at least 20 times per year (the duration of an observation/demonstration, including feedback, varied from 40 to 75 minutes).

The ADOR-IIR experimental group: Implementing the IIR program with pupils with the weakest reading skills in the ADOR program in first-grade

The ADOR program in grade 1 is accompanied by the IIR remedial program with the goal to recover pupils with the weakest reading skills. As mentioned above, the program is a translation and adaptation of the DIR reading support program (see Boyer, 2012, for a more detailed description), an existing French-language program in Québec. The IIR program for Grade 1 rolls out in the winter or spring, and targets the first graders with the weakest reading skills who also participate in the ADOR program in their regular classroom. These children are identified by measuring their oral reading of a text that they have not read before. Another measurement, of comprehension and reasoning on silent reading, is made using one or more small texts that the children have never read before. Then, the school's 8-12 lowest-performing first graders are selected, regardless of other criteria (e.g., hyperactive, dyslexic, behaviour problems, language at home, or

 ISBN : 978-2-923805-66-5

having cooperative parents).

Like its French-language counterpart, IIR is an explicit and direct reading instruction program. This remedial program rolls out over an 8-10-week period, with 2 consecutive hours of intervention per day. In all, the IIR program must include a minimum of 76 hours of intervention. The basis of the IIR program, which is applied by the school's special-aid teachers, includes activities from ADOR program, with some special features for students who are at-risk or have learning disabilities. In addition, certain empirical data specific to remedial intervention are incorporated into the DIR program (Boyer, 2012). The content of the intervention is similar to the activities from the regular classroom, but the pace is faster, the instructional materials are different, and some of the activities and specific classroom-management techniques are adapted. The program aims to reduce the gap between IRR students and the other pupils in the regular classroom.

During the school year, IRR special-aid teachers participated in eight days of training led by the school board's trained coaches. The coaches followed up in the classroom (with the pupils and their teacher) to demonstrate instructional activities and classroom management techniques and observe the teachers' level of application. Observation grids were used and special-aid teachers received feedback. Each remedial teacher was observed at least 12 times per year (an observation/demonstration including feedback lasted a full day).

The control group: Use of the Québec Ministère de l'éducation curricula in regular classes with all first-grade pupils

The control-group classes followed the Québec Ministère de l'éducation's regular curricula (2001), which was monitored by the school board's pedagogical services staff. The *Whole Language* and constructivist approaches strongly influenced the pedagogical services staff's discourse and recommendations for how the curriculum must be applied. In short, the recommended approach to reading instruction is to read large texts (from big books or children's literature) every day to the group, to guide reading with questions, to have pupils globalize specific words, to emphasize comprehension and sense-making in reading, and to verbalize words using anticipation, often based on the illustration, title, context, and other words in the sentence or text. Ideally, these tasks are done as part of a "meaningful and authentic" educational project based on student interests that leads to some form of production (writing, drawing, etc.). Writing and reading are closely integrated into these projects and informal formative assessments (student interviews, work samples, etc.) are preferred. The creation of a portfolio (student work samples) reflecting the student's development in reading and writing is encouraged. Decoding (phonics) is not prohibited, but it is recommended that it be done briefly, only when needed and in response to difficulties in reading.

Monitoring of the application of this pedagogy is ensured through various trainings offered by the school board's pedagogical services staff whom are responsive to teachers' requests and usually provide training that can lead to follow-up in the classroom. However, we do not know how many training sessions and follow-ups have been offered and what kind of feedback was used.

The control group: Special-aid teachers using a variety of remedial interventions with the weakest pupils in reading skills in first-grade regular classes which uses the Québec Ministère de l'éducation curriculum

All classes in the control group received remedial instruction. Usually, home-teachers select pupils with weak skills to receive remedial services from special-aid teachers, and the process is monitored by the principal. Delivery of special-aid services vary. There is the "pull-out," students are taken out of class 1-3 times a week for 20-40 minutes sessions, depending on the child's difficulties and the special-aid teacher's availability. Students may be seen individually or in small groups. The remedial progam is similar to regular class content and activities, sometimes using the same materials and sometimes using different materials. Some special-aid teachers may use also more play-based activities in their interventions. Others provide activities that complement the learning in the regular classroom. Some schools ask special-aid teachers to work exclusively or partially right in the regular classroom. Usually the special-aid teacher working in the regular classroom, helps the teacher by working with the weaker pupils. Sometimes the home-teacher

 ISBN : 978-2-923805-66-5

and special-aid teacher work as a team, sharing tasks and pupils.

Supervision of remedial services is provided by the school board and its educational consultants. The latter usually respond to the requests of special-aid teachers by providing materials, training, or support. We do not know how much training and follow-up was provided to special-aid teachers or what type of feedback they used.

Standardized measurement instruments used by Savage (2005), and Savage and Deault (2007)

The Group Reading Assessment and Diagnostic Evaluation (GRADE) is a standardized reading and listening-comprehension test has Canadian norms. GRADE has proven to be reliable and valid for measuring reading and listening skills. It can be administered to an entire class of young children.

Pretests

The GRADE's Word reading, Word meaning, and Listening comprehension subtests were administered during the pretest between September 15 and October 15, in both 2004 and 2006. These subtests were administered in the classroom to all the children at once for both the experimental and control groups.

In the Word Reading subtest, children are asked to check one of the four words per item. The word is first read aloud by an examiner, then used in a sentence, and finally repeated. In the Word Meaning subtest, children are asked to read a word and select one of the four pictures that best depicts its meaning. In the Listening Comprehension subtest, children are asked to listen to a short sentence or excerpt and to select one of the four pictures that best depicts the meaning of the text they just heard.

Posttests

Posttests were administered between April 15 and the end of June 2005 and in May 2007. The Listening Comprehension, Word Reading, and Word Meaning subtests were again used. Two additional GRADE subtests, Sentence Comprehension and Reading Comprehension of text passages, were added. For the Sentence Reading Comprehension subtest, students are asked to read a sentence silently and select the word (out of four) that best matches the sentence. For the Passage Comprehension subtest, students silently read a short passage of three to four sentences. They then read a question and select an answer among four choices.

Hawthorne effect

To control for the Hawthorne effect, the arithmetic subtest of the *Wide Range Achievement Test* (WRAT) was administered. Its first 10 items were presented to pupils at both the pretest and the posttest.

Additional measures

Additional qualitative assessments were also conducted. The researchers interviewed teachers and observed the classrooms, using the Atmosphere, Instruction, Management, Student Engagement (AIMS) test developed by Pressley et al. 2001. In addition, pupils completed the Classroom Environment Scale (CES; Fraser 1986), which measures their perceptions of the classroom. The results of these additional measures were correlated with student progress in reading.

Hypotheses[6]

Since ADOR and IIR are partly based on Explicit and Direct Instruction frameworks and build on the evidence for learning to read, we postulated the following:

1. The ADOR-IIR group is expected to outperform the control group on direct measures of reading learning in first-grade, particularly in Word reading, Sentence reading comprehension, and Passage reading comprehension.

Since learning to read and reading itself can be influenced by oral language development (Lervåg, Hulme, & Melby-Lervåg, 2018):

2. It is expected that the ADOR-IIR group will outperform the control group on measures of Word meaning and Listening comprehension.

Since reading is not very solicited in first-grade math and even less so in arithmetic:

3. The ADOR-IIR group should perform similarly to the control group on the arithmetic measure.

 ISBN : 978-2-923805-66-5

Results

In this study, missing data accounted for less than 1 % of all data and were caused by student non-attendance at assessments. Prior to conducting the statistical analyses, the external research team verified that the data were normally distributed. Stanines ranging from 1 to 9 with a mean of 5 were used for this analysis. The analyses show that there was no significant kurtosis or skew in the data ($s < 1$ns, and $s < 1$ns). However, the pretest results for the arithmetic subtest showed positive skewness. This variable was normalized by square-root transformation.

Table 1 shows the means and standard deviations for each of the reading subtests in the pretests and posttests for the ADOR-IIR experimental group and for the control group. An effect size is also provided for each measure. There were no significant differences between groups on the pretests, except for the Listening subtest ($F = 8{,}392$, $p < 0.05$). This difference was accounted for in the analyses performed subsequently (ANOVA). The posttest results for each of the reading subtests showed that the pupils who received the ADOR-IIR program had higher scores than students in the control group. The Word reading subtest showed a significant difference ($p < 0.05$) to the advantage of the ADOR IIR group with an effect size of 0.20. The difference in the Word meaning subtest was also significant ($p < 0.001$) in favour of the ADOR-IIR group with an effect size of 0.46. The Listening comprehension subtest showed a significant difference ($p < 0.01$) in favour of the ADOR-IIR group, with an effect size of 0.29.

In addition, the largest differences between the two groups are observed in Sentence comprehension ($d = 0.59$; $p < 0.01$) and Passage comprehension ($d = 0.84$; $p < 0.001$). These two tests measure one of the ultimate goals of reading: comprehension[7].

The results are not significant for the arithmetic test.

Hypotheses 1 and 2 were confirmed, as the ADOR-IIR group outperformed the control group on all relevant measures. The ADOR-IIR group performed better than the control group on the Word reading, Sentence reading comprehension, and Passage reading Comprehension subtests (Hypothesis 1), and these results were all significant ($p < 0.05$ to $p < 0.001$). The measures of the Listening comprehension and Word meaning subtests all showed a significant difference ($p < 0.01$ to $p < 0.001$) favouring the ADOR-IIR experimental group over the control group.

Hypothesis 3, which tests for the Hawthorne effect, was also confirmed. Since the result of this comparison was not significant—the two groups performed approximately the same—the Hawthorne effect was not at play.

Of the supplementary measures used, the only one that correlated with pupils' reading progress was the CES questionnaire that the students completed. Since the results presented by the external researchers did not allow us to compare the perceptions of pupils in the ADOR-IIR group with the perceptions of control group students, we consider it irrelevant to present the results.

Discussion

This study shows that, when used in combination, the ADOR and IIR programs result in superior reading performance as measured by subtests of the standardized GRADE instrument. All observed differences were statistically significant. To understand the impact of the effects of ADOR and IIR on first graders' reading achievement, we will now compare these effect sizes (Cohen's *d* effect sizes) in terms of learning gains.

Standardized tests of academic achievement in the United States show that, on average, one year of reading instruction for grade 1 pupils is equivalent to $d = 0.97$ (Hill, Bloom, Black, & Lipsey, 2008). If this is also true for the GRADE test and its subtests, theoretically, the number of months of learning gains implied by the effect sizes reported in Table 1 can be calculated. For example, the effect size of 0.20 for Word reading would correspond to a learning gain of approximately two months (0.20/0.97 = 0.21 of a 10-month school year, i.e. two months[8]). For the Word meaning measure, ADOR-IIR would gain almost five months over the control group (0.46/0.97 = 0.47) and three months more learning on the Listening comprehension measure (0.29/0.97 = 0.30). It is important to note that the ADOR-IIR group performed significantly better than the control group on reading comprehension

 ISBN : 978-2-923805-66-5

Table 1. Combined pretest and posttest scores in reading and in 2004–2005 and 2006–2007.

Group	Pretest	Posttest	R max	Effect size (*d*)
Word reading				
ADOR and IIR group	3.72 (1.60)	6.00 (2.19)	9	0.20
Control group and remedial services	3.46 (1.57)	5.44 (2.32)		
Word meaning				
ADOR and IIR group	3.71 (1.77)	5.78 (1.95)	9	0.46
Control group and remedial services	3.36 (1.79)	4.90 (2.19)		
Listening Compréhension				
ADOR and IIR group	4.56 (2.04)	5.62 (1.96)	9	0.29
Control group and remedial services	4.14 (1.79)	5.01 (1.90)		
Sentence Compréhension				
ADOR and IIR group	Nil	5.13 (2.29)	9	0.59
Control group and remedial services		4.12 (2.30)		
Passage comprehension				
ADOR and IIR group	Nil	5.05 (2.07)	9	0.84
Control group and remedial services		3.57 (2.12)		

Note 1. Standard deviations are shown in parentheses.

Note 2. The effect sizes shown are calculated according to Cohen's *d* and the $\eta 2$ were transformed with the formula $d = 2 \times \sqrt{\frac{\eta^2}{1-\eta^2}}$.

Note 3. R Max represents the maximum stanine for the reading subtests.

 ISBN : 978-2-923805-66-5

measures. Two GRADE subtests measure Sentence reading comprehension, which would represent a six-month learning gain over the control group (0.59/0.97 = 0.61) and Passage reading comprehension, which would show an extraordinary learning gain of nearly nine months (0.84/0.97 = 0.87). In other words, on this measure, pupils in the ADOR-IIR group would have completed the equivalent of two school-years of learning in only one school year. Although these conversions of effect sizes to months of learning are only theorical[9] they suggest that the ADOR-IIR programs can have significant effects, particularly with respect to promoting comprehension, the ultimate goals of reading instruction.

In the next section, we take a critical look at the methodology of Savage and Deault's (2007) study.

Some of the elements used to develop Best Evidence

For several years, Slavin and his team have been proposing, as part of their *Best Evidence* concept, a consideration of certain methodological elements as a way to identify and select the best experimental research for conducting valid meta-analysis (Cheung & Slavin, 2016). The reason for this is quite simple: the methodology of experimental research correlates with the effects obtained (Cheung & Slavin, 2016). We will apply some of Slavin et al.'s findings to Savage and Deault's (2007) research, in the goal of assessing how close this research is to the *Best Evidence* concept[10].

The use of a standardized or an independent[11] measurement is clearly preferable to a "homemade" device

The choice to use a standardized or an independent measurement, vs a "homemade" device, can strongly influence effect size. In their analysis of seven mathematics research studies, Slavin and Madden (2011) reported that the effect size for homemade measurements is 0.45 and -0.03 for standardized measures. Cheung and Slavin (2016) found an effect size of 0.40 for homemade measures and of 0.20 for standardized measures in a meta-analysis of 645 studies on learning in reading, mathematics, and science from pre-K through 12th grade.

The study on which this paper is based (Savage & Deault, 2007) used a measure that is independent of the program designer and study researchers. Subtests from the GRADE instrument, standardized to a Canadian English-language sample, were used to measure student performance.

The sample size of a study has a major influence on the observed effect size

Slavin and Smith (2009) found that effect sizes vary with the number of subjects in the study. Studies with fewer than 51 subjects obtained an effect size of 0.44, while studies with 51—100 subjects obtain an effect size of 0.29. When the sample size is larger than 2,000 subjects, the effect size is only 0.09. Pellegrini (2017; see Slavin, 2018) obtained an effect size of 0.37 with multi-field pedagogical research with 60 or fewer subjects and an effect size of 0.13 when the sample was greater than 250. In a meta-analysis of studies on mathematics in elementary and secondary, Slavin and Smith (2009) found that a sample size of 50 or fewer subjects yielded an effect size up to 3.67 larger than a sample of 401 to 1,000 participants (an average effect size of 0.44 for studies with 50 subjects or fewer, and an average effect size of 0.12 for studies with 401 to 1,000 participants).

The study we draw on here, by Savage and Deault (2007), had a sample size of 480 pupils (204 students in the ADOR-IIR experimental group and 276 in the control group), therefore reducing the possibility that the observed results are overinflated by a small sample.

The sample attrition should not exceed 15 %

Neitzel, Lake, Pellegrini, and Slavin (submitted) noted that studies with attrition rates greater than 15 % should be excluded from meta-analysis so as not to affect the observed results. This suggestion can also be found in the *What Works Clearinghouse* (2020) recommendations of standards for conducting meta-analyses.

The study by Savage and Deault (2007) had a very low attrition rate of less than 1 % missing data.

ISBN : 978-2-923805-66-5

The experiment must be at least 12 weeks between pretest and posttest

Cheung and Slavin (2016) recommend that studies span at least 12 weeks to reflect the effects of common practices over a school year. The shortest interventions (less than 10 instructional hours; see, for example, Wanzek et al., 2017) appear to produce the highest effects which are artificial.

Again, Savage and Deault's (2007) study measured the effects of a full school year on two occasions. Consequently, their study reflects the effect of a possible normal practices over a school-year.

Experiments should randomly assign students to the experimental and control groups (experimental method) or statistical adjust between groups (quasi-experimental method)

Though random assignment is an important feature of the scientific method (Cheung & Slavin, 2016), it is rarely used in educational research. Savage and Deault's (2007) study does not use random assignment, but it did make the necessary statistical adjustments.

In summary, Savage and Deault's (2007) study performs very well on many of the important elements underlying Slavin's Best Evidence concept and should have been retained in a meta-analysis made by Slavin and his collegues.

Limitations of the study

This study has some limitations. One is that the respective effects of the ADOR program and the IIR program could not be determined, since the two programs were not measured separately. However, this does not invalidate the observed results; it simply limits a more detailed understanding of each program's contribution to the results.

Although not mentioned in Savage and Deault (2007), fidelity in implementing the ADOR and IIR programs[12] was ensured by the coaches[13] in both years of the experiment. In contrast, the fidelity of the program's implementation in the regular classes and remedial programs was not monitored as closely or as intensively as the programs in the experimental group. Although we can demonstrate that the ADOR and IIR programs were indeed effectively implemented, we have no information on the extent to which the programs of the Québec Ministère de l'éducation (regular curriculum) were implemented in the control group. This weakness is widespread in educational research, and Savage and Deault's (2007) study is no exception.

Conclusion

We reiterate that Savage and Deault's (2007) study meets many of the elements of Slavin's *Best Evidence* concept which he used to identify the most rigorous research to conduct high-quality scientific meta-analyses.

Combined, two first-grade reading programs—one intended for students in regular classrooms and another for remedial instruction—based on Explicit instruction, Direct instruction, and other largely evidence-based elements demonstrated effectiveness superior to the Québec Ministère de l'éducation's curriculum in the regular classroom and remedial programs on all measures made in an English-speaking school board in Québec. The largest effects were on subtests measuring reading comprehension. An effect on Passage comprehension was observed that could theoretically result in a gain in learning the equivalent of an additional school-year. As other researchers (National Reading Panel, 2000; Bissonnette, Gauthier, Richard, & Bouchard, 2010; Stockard, Wood, Coughlin, & Rasplica Khoury, 2018) have shown, our study confirms that programs based on Explicit and Direct Instruction could promote better learning compared to other programs.

 ISBN : 978-2-923805-66-5

References

Bissonnette, S., Richard, M., Gauthier, C., Bouchard, C. (2010). Quelles sont les stratégies d'enseignement efficaces favorisant les apprentissages fondamentaux auprès des élèves en difficulté de niveau élémentaire ? Résultats d'une méga-analyse. *Revue de recherche appliquée sur l'apprentissage*, 3(1), 1-35.

Boyer, C. (1993). *L'enseignement explicite de la compréhension en lecture : modèles d'activités d'enseignement*. Boucherville : Graficor.

Boyer, C. (2012). *DIR: A Remedial Program for Reading — Intensive Intervention in Reading*. Montréal : Éditions de l'Apprentissage. https://www.editionsdelapprentissage.com/index.html

Boyer, C., Bissonnette, S. (2021). Comment exercer une gestion rationnelle axée sur les résultats ? Exemple de la mesure de l'effet d'un programme orthopédagogique sur le rendement scolaire des élèves. *Enfance en difficulté*, 8, 97-128.

Carnine, D., Silbert, J., Kameenui, E. J., Tarver, S. G. (1997). *Direct instruction reading*. Columbus, OH: Merrill.

Cheung, A. C., Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283-292.

Dion, É. (2012). Compte rendu de [Boyer, C. (2010). Le programme orthopédagogique DIR en lecture. Montréal, Qc : Éditions de l'apprentissage]. *Revue de psychoéducation*, *41*(2), 283–286.

Forget, J. (2012). Les contributions de la pédagogie comportementale à l'éducation spécialisée. *Enfance en difficulté*, *1*, 7–45.

Fraser, B. J. (1986). *Classroom environment*. London: Croon Helm.

Hill, C. J., Bloom, H. S., Black, A. R., Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, 2(3), 172-177.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. https://dspace.library.uu.nl/bitstream/handle/1874/16884/kirschner_02_cognitive_load_theory_implications_cognitive.pdf?sequence=1

Lervåg, A., Hulme, C., Melby Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821-1838.

Little, S. G., Akin-Little, K. A. (2003). Classroom management. Dans O'Donohue, W., Fisher, J., et Hayes, S. (Éditeurs), *Empirically supported techniques of cognitive behavioral tehrapy : A step-by-step guide for clinicians* (65-70). New York: Wiley.

Ministère de l'Éducation du Québec (2001). Programme de formation de l'école québécoise. Éducation préscolaire, Enseignement primaire. Québec : Gouvernement du Québec.

National Reading Panel (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. National Institute of Child Health and Human Development, National Institutes of Health.

Neitzel, A., Lake, C., Pellegrini, M., Slavin, R. (soumis). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*.

Oakhill, J., Cain K., Elbro, C. (2015). *Understranding and Teaching Reading Comprehension—A handbook*. Oxon: Routledge.

Pressley, M., Wharton-Mc Donald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, 5,35-58.

 ISBN : 978-2-923805-66-5

Savage, R. (2005). An independent evaluation of the ADOR early reading program and other forms of effective teaching of reading in Level 1 Cycle 1. Montréal : Université McGill.

Savage, R., Deault, L. (2007). An independent evaluation of the ADOR early reading program and other forms of effective teaching of reading in Level 1 Cycle 1 2004-2007. Montréal : Université McGill

Seligman, M. E. P. (1995). *The optimist child. A Proven Program to Safeguard Children Against Depression and Build*. New York : Houghton Mifflin Harcourt.

Slavin, R. (2018). Small studies, big problems. Repéré à https://robertslavinsblog.wordpress.com/2018/09/13/small-studies-big-problems/

Slavin, R., Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness, 4*(4), 370-380.

Slavin, R., Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational evaluation and policy analysis, 31*(4), 500-506.

Stockard, J., Wood, T. W., Coughlin, C., Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half-century of research. *Review of Educational Research, 88*(4), 479-507.

Stuart, M., Stainthorp, R. (2016). *Reading Development and Teaching*. Sage Publications.

Wanzek, J., Petscher, Y., Al Otaiba, S. A., Rivas, B. K., Jones, F. G., Kent, S. C., Schatschneider, C., Mehta, P. (2017). Effects of a year-long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology, 109*(8), 1103–1119.

What Works Clearinghouse (2020). *Procedures handbook (Version 4.1)*. Institute of Education Sciences, US Department of Education.

Willingham, D. T. (2017). *The Reading Mind : A Cognitive Approach to Understanding How the Mind Reads*. San Francisco : Jossey-Bass—A Wiley Brand.

 ISBN : 978-2-923805-66-5

Notes

1. The director general of the school board gave its permission for the authors to use the reports results in this writing.

2. The designer of ADOR and IIR, Christian Boyer, is one of the authors of this text.

3. The *Group Reading Assessment and Diagnostic Evaluation* (GRADE™) is a diagnostic reading test that determines the achieved level of reading and listening skills of pupils in grades K-12.

4. The Hawthorne effect refers to the effect of simply participating in an experiment, which tends to lead to greater motivation among subjects in the experimental group compared to subjects in the control group, which may affect the results of the experimental group.

5. Exact rate is the number of words students read correctly in one minute without help when reading a text out loud that they have never seen before.

6. These are the authors' (Bissonnette, Boyer, & Morneau-Guérin, 2022) hypotheses and not those of Savage and Deault (2007).

7. The effect sizes of the 2004—2005 report (Savage, 2005) on the same measurements are $d = 0.59$ ($p < 0.001$) and $d = 0.87$ ($p < 0.001$), respectively.

8. In other words, pupils in the ADOR-IIR group would have gained the equivalent of two months more learning than the control group.

9. This analysis is intended to illustrate the significance of the results obtained.

10. Slavin's *Best Evidence* concept considers certain methodological elements when selecting research for meta-analysis and analyzes these elements as variables moderating effect size.

11. It is preferable to use a third-party measurement, i.e. one that was developed independently of the study's designers or researchers, to one developed by the designers and researchers themselves.

12. With the lost data, according to Christian Boyer's recollection, the application fidelity of the programs varied from 50 % to 98 % for ADOR with an average of 79 % and from 40 % to 98 % with an average of 71 % for IIR.

13. The trainers were trained and supervised by Christian Boyer.