FUNDACIÓ JAUME BOFILL

ivàlua | Institut Català d'Avaluació de Polítiques Públiques

What Works in Education?

Using Evidence to Improve Education

2

November, 2015

Are individual tutoring programs effective in addressing diversity?

Miquel Àngel Alegre Canosa

What grouping strategies respond to criteria of efficiency and equality?

Gerard Ferrer-Esteban

"For far too long education has been based on inertia and traditions with changes implemented hinging on intuition or unfounded beliefs. The "*what works*" movement takes the world of education by storm with one clear objective: to promote evidenced-based educational policy and practices. Ivàlua, together with the Jaume Bofill Foundation join forces to promote this movement here in Catalonia."

Fundació Jaume Bofill

ivàlua | Institut Català d'Avaluació de Polítiques Públiques

What Works in Education?

Using Evidence to Improve Education

Are individual tutoring programs effective in addressing diversity?

Miquel Àngel Alegre Canosa

Analyst at the Catalan Institute for the Evaluation of Public Policy (Ivàlua). Expert in evaluation of education policy and author for the *What Works in Education* report.

Motivation

The initiatives discussed in this article share a commitment to an individual tutoring system one-to-one (one student, one tutor), as a mechanism for addressing diversity and improving cognitive and non-cognitive skills of students with academic and, very often, social disadvantages. These interventions, which we will refer to as individual tutoring programs (ITP), are diverse in nature. However, in this article we will focus on three different ITP methods:

- Tutoring intervention.
- School-based mentoring.
- Peer tutoring.

These ITPs have been implemented within the Catalan education system in a disproportionate manner. While tutoring intervention or school-based mentoring programs are currently subject to limited implementation, peer tutoring is beginning to gain ground in educational circles, especially in primary education centers.

However, it cannot be said that the deployment of the different ITPs responds to evidence-based criteria concerning their greater or lesser efficiency, or even less still to cost-efficiency and cost-benefit ratio. The fact of the matter is that nowadays we enjoy access to a wealth of sound evidence concerning the impact of these programs.

Even though the vast majority of this evidence is derived from evaluation and reviews of studies concerning programs developed in the United States and the United Kingdom, a review of the data provides us with knowledge which might be useful when it comes to evaluating, either positively or negatively, any commitment to ITPs as a tool for addressing educational diversity and equal opportunities here in Catalonia.

This review will deal with three different methods of individual tutoring programs (ITP): tutoring intervention, school-based mentoring and peer tutoring.

Questions influencing the review

The focus of this review encompasses three types of ITP, each with its own function, theory of change and stakeholder profiles (students and tutors): tutoring intervention, mentoring programs and peer tutoring. Table 1 provides the principal characteristics of each system and includes examples of programs implemented here in Catalonia for each category.

Table 1.
Individual tutoring programs. Types and characteristics

Individual tutoring programs	Prominent function (in relation to recipient students)	Theory of change (basic mechanism)	Profile of recipient students	Tutor profile	Examples in Catalonia
Tutoring intervention	Improve cognitive skills (academic circles)	Intensive school intervention work and study planning	Students with academic issues (children & adolescents)	Specialist or specific know-how (teachers or volunteers)	• Centers own tutoring & intervention school initiatives • Lexcit Program (Jaume Bofill Foundation & Obra Social "laCaixa") • Learning Programs *Servei de format 1:1*
School-based mentoring	Improve cognitive and noncognitive skills (attitudes & motivations)	Significant relationship and establishing positive model	Students with academic issues and social problems (children & adolescents)	Young or adult volunteers with mixed profiles (according to positive model chosen)	• Rossinyol Program (UdG, Servei Solidari & KM 0) • *En Tàndem* Program (AFEV) • Municipal educational support programs
Peer tutoring	Improve cognitive and non-cognitive skills (attitudes and aptitudes)	Cooperation and autonomous work	Mixed students (ages & academic levels)	Students from same center (from the same class or higher level courses)	• Programs developed by centers during school hours (reading buddies, maths buddies, students-guide, etc.)

Source: Miquel Àngel Alegre Canosa.

Considering the diversity and particular features of ITPs, we propose answering the following questions: Are ITPs effective in improving academic outcomes of recipient students? And, what about improving students' non-cognitive skills (in social, emotional and attitudinal outcomes)? Which ITP systems have demonstrated greater efficiency in terms of improving academic performance and attitudes? Which group of students reap greater rewards and which reap less from the different ITP models? And, finally, in the proposed plan: to what extent is it recommendable to expand the application of the different ITP systems in Catalonia?

Reviewing the evidence

This study is based on data compiled from a total of sixteen meta-analyses and three non-systematic reviews of the efficiency of ITP implemented in different countries, above all in the United States and the United Kingdom. In this sense, what we set out to present in this article is what we refer to as a "review of reviews". In line with the "what works" perspective, we will focus on those reviews and meta-analyses which select only extremely stringent methodological impact evaluations as their focus of attention, prioritizing those that concentrate on studies of an experimental nature. Complementing this process, we will also analyze the arguments put forward with reference to evaluations of experimental pilot programs of particular relevance.

This review is organized based on the three ITP models under consideration.

Does tutoring intervention work?

We will begin by discussing the ITP model which is clearly associated with the academic environment. One-to-one tutoring intervention is designed to address students with specific issues in certain skills areas and focuses on improving academic performance. These tutoring sessions tend to be deployed within the education center itself, even though they may take place outside school hours, and usually consist of two sessions weekly of thirty to forty-five minutes' dosage throughout the greater part of the academic year.

Programs that incorporate tutoring intervention however, are diverse, depending on elements, such as: tutor profile (specialist teachers or not, teacher's assistants, volunteers, etc.), the characteristics of students (age, type of educational deficit, associated social problems, etc.), frequency, location and time when tutoring is delivered, subjects taught, complementary initiatives (for example, training activities for tutors), etc.

Some of these programs have been evaluated experimentally. For example, in the United States, in the area of tutoring intervention in reading, we can reference experimental studies such as the Reading Recovery [1][2], Reading Rescue [3], Experience Corps [4][5], Reading Partners [6] and Sound Partners programs [7] [8]. In the United Kingdom we can find the experimental studies, Time to Read [9][10], TextNow Transition Programme [11], Switch-on Reading [12] and Catch Up Literacy [13].

In the area of tutoring intervention in maths, there is a much smaller volume of empirical evidence available. In this case we can mention experimental evaluations of the Catch Up Numeracy [14] and Numbers Count programs [15][16] in the United Kingdom, and the West Philadelphia Tutoring Project [17] and Match Tutoring Model [18] in the United States. It is not surprising therefore that the principal meta-analysis synthesis of evidence concerning the effectiveness of tutoring intervention concentrates on the area of reading skills (please see Table 2.

Box 1.
Time to Read (Northern Ireland)

In 2015, Time to Read was implemented in 100 primary schools in Northern Ireland, an intervention affecting around 1,000 students between the ages of 8 and 10 years with reading comprehension deficit. Volunteer mentors were paired up with students and followed a one-to-one tutoring system throughout the entire school year. Up until 2008, the 30-minute tutoring sessions were held once a week, outside the reference class group.

An initial experimental pilot program, carried out between September 2006 and June 2008, produced disappointing outcomes in the majority of the areas measured (basically, variables related to reading comprehension, attitudes towards schooling and self-esteem). After 2008, and as indicated by the same evaluation report, the program proceeded to double the frequency of tutorials, delivering two 30 minute sessions weekly.

Between October 2008 and June 2010, the second experimental evaluation of the program was carried out. The study was based on a sample of 512 students in 50 primary schools. The procedure consisted of randomly assigning, within each school and among students prioritized by the teacher, those students who were included in the program (263 students) and those who were excluded (249 students), in an effort to compare the corresponding outcomes after a two-year follow-up period. The outcomes from this second evaluation enabled reviewers to attribute to the program a positive impact on outcomes, such as decoding capacity, speed, and fluency in reading. On the other hand, the program did not appear to be effective in the area of reading comprehension or when it came to increasing students' enjoyment of reading.

For further information:
Miller, S., Connolly, P. (2012). "A Randomized Controlled Trial Evaluation of Time To Read, a Volunteer Tutoring Program for 8- to 9-Year-Olds". *Educational Evaluation and Policy Analysis*, 35(1), 23-37.

Miller, S., Connolly, P., Maguire, L. K. (2012). "The Effects of a Volunteer Mentoring Programme on Reading Outcomes Among Eight-To Nine-Year-Old Children: A Follow Up Randomized Controlled Trial". *Journal of Early Childhood Research*, 10(2), 134-144.

One-to-one tutoring intervention is designed to address students with specific issues in certain skills areas and focuses on improving academic performance. These sessions tend to be held outside regular school hours and consist of two sessions weekly throughout most of the school year.

Table 2.
Tutoring intervention. Meta-analyses reviewed

Meta-analysis (reference country)	Number of studies included	Skills considered	Profile of students in programs	Profile of tutors in programs	Number of sesions	Measurement of effect*
D'Agostino & Murphy (2004) [19] (United States)	36	Reading	First year primary school students	Specialist teachers	30 minute sessions 5 days per week for between 12-20 weeks	+0.32
Elbaum *et al.* (2000) [20] (United States)	29	Reading	Primary students (different years)	Mixed profiles	30 minute sessions 5 days per week Over 1 year (average)	+0.41
Jun *et al.* (2010) [21] (United States)	12	English (reading, writing and vocabulary)	Secondary students (12-18 years.)	Mixed profiles	7 hours or less up to 16 hours or more	+0.70 (with adult tutors)
Ritter *et al.* (2009) [22] (United States)	21	Reading, writing and maths	Primary students & junior secondary	Non-professional adults (volunteers)	From one 60 minute session per week for a month; up to two 30 minute weekly sessions for two years	+0.30 (reading) +0.27 (maths)
Slavin *et al.* (2011) [23] (United States)	97 (total) 38 (1:1)	Reading	Primary students (different years)	Mixed profiles	Include programs (one-to-one & others) Lasting over 12 weeks	+0.62

Source: Prepared by authors from https://educationendowmentfoundation.org.uk/uploads/pdf/One_to_one_tuition_Technical_Appendix.pdf
* The standardized effect value is given, in accordance with Cohen's effect size measure (1988) [24]. In this way the measure of impact must be compared between programs. Based on Cohen's indications, the following is generally true: values similar to or lesser than 0.2 indicate a small effect size; values similar to 0.5, a medium effect size; values in the region of or greater than 0.8, a large effect size.

What does the empirical evidence tell us about the effectiveness of one-to-one tutoring intervention?

In general, the data shows that one-to-one tutoring intervention can produce a significant positive impact on students' learning process. In accordance with the summary of evidence from the Education Endowment Foundation, this impact capacity accelerates student academic course learning by five months, on average. We should point out at this stage that the observation window for the studies reviewed tends to be quite reduced (normally, this window closes just after the intervention ends), a fact that makes formulating a clear hypothesis concerning the possible impact of these programs in the medium to long-term difficult.

Although, conclusions from meta-analyses of the evaluations reviewed suggest that certain intervention strategies work better than others depending on the context and depending on the target population of the intervention. In particular:

- **Skills covered.** Effectiveness of tutoring intervention seems principally contrasted in the area of reading skills. In the field of mathematics, at present there is less accumulated evidence available, even though some recent experimental pilot studies also show promising outcomes in this area [14][15][16].
- **Program dosage.** Tutoring intervention programs work better when the dosage of the sessions, frequency and extension adapt to the students' learning requirements [23]. Some studies conclude that the more intensive programs (in terms of the number of sessions per week) and with an overall dosage of between three to five months, tend to have a greater effect [20][21].
- **Type of pairing.** We should point out that it is of great importance that the pairing between student and tutor remains stable, thereby guaranteeing the establishment of significant bonds throughout the intervention [25]. There is no evidence regarding the possible added value of paired tutors and students based on same sex or same ethnic origin.
- **Structure of contents.** It is clear from the evidence that intervention content programming should follow a structured curriculum focusing on the different aspects involved in the skills area covered by the intervention program. In the area of reading comprehension, tutoring intervention has been shown to be particularly effective in improving the more basic functions of the skill (decoding, speed and fluency in reading), and not so much in terms of reading comprehension [10] [22]. In any event, it appears clear that these tutoring sessions work better when they are linked to, and supplemented by, the regular dynamics of the schooling process (standard reference curriculum, group, class and teacher), both in terms of content programming as well as in the definition of procedures [23].
- **Tutor profile.** Tutoring delivered by qualified teachers tends to work better and to produce a better cost-gain balance that those performed by teachers' assistants or volunteers [26][27]. Nevertheless, the effectiveness of the latter increase when assistants and volunteers are provided with suitable training and are subject to valid monitoring by specialist teachers [20][23].

- **Characteristics of treatment students.** In relation to academic level, the vast majority of the studies evaluate programs focusing on students within a similar age group (at the beginning or midway through primary education); however, we can observe that programs focusing on secondary students have not proved to be effective [12][13][18]. On the other hand, the majority of the studies reviewed conclude that the effectiveness of tutoring intervention does not depend on the pre-intervention academic level of the students or their sociodemographic characteristics (sex, ethnic group and socioeconomic background).

One-to-one tutoring intervention can produce a significant positive impact on students' learning process. This impact capacity accelerates student academic course learning by five months, on average.

Do school-based mentoring programs work?

In general, a mentoring relationship is that which is established between an individual with a certain level of personal and socioprofessional experience (mentor) and a person of a younger age who is deemed to be vulnerable in some way or to be at risk (mentored). Through this relationship, the mentor offers a positive individualized model and assists the mentored student in improving their attitudinal, skills and emotional outcomes.

Here we will focus on what are referred to as school-based mentoring programs. [1] These programs target school-going age groups and focus on both academic as well as attitudinal content. They are frequently delivered in the same education center, during the school year, outside school hours and usually once a week.

School-based mentoring programs have undergone very significant growth in recent years, principally in the United States. However, within the category of school-based mentoring, there are a wide range of different intervention programs and strategies, which vary between one another depending on the following: mentor profile (professionals, adult volunteers, university students, elderly individuals, etc.), mentored student's profile (age, associated social problems, educational deficiencies, etc.), skills area targeted (cognitive or non-cognitive), dosage of sessions, etc.

There are many school-based mentoring programs that have been evaluated experimentally. In the United States, evaluations which stand out include the Big Brothers Big Sisters of America program [29] [30], the Quantum Opportunity Program [31], the SMILE (Study of Mentoring in the Learning Environment) program [32] and the Student Mentoring Program [33].

School-based mentoring programs are designed to target school-going age groups and focus on both academic as well as attitudinal content. Ordinarily, mentor and mentored tend to meet throughout the whole school year once a week.

1 This kind of mentoring model differs from community-based mentoring, which focuses on the set of aspects which effect young people's lives and their transition into adulthood (in training, occupational, family, residential, relationships and health-related issues, etc.) [28].

Box 2.

Big Brothers Big Sisters of America (BBBSA) (United States)

BBBSA is the longest standing and largest mentoring program in the United States. The BBBSA organization was set up in 1977 and came about as a result of the merging of two organizations; Big Brothers and Big Sisters International. Traditionally the program focused on community-based mentoring but after 2000, the BBBSA program began prioritizing its focus and dedication towards school-based mentoring.

BBBSA pairs vulnerable boys and girls between the ages of 6 and 18 years of age with volunteer adults within the framework of a one-to-one relationship offering support and confidence. The aim of this relationship is to achieve a series of impacts on the mentored student in emotional, attitudinal, social and academic outcomes.

In its school-based aspect, the BBBSA was piloted and experimentally evaluated between 2004 and 2005. 1,139 primary and secondary school students took part in the study (all of whom were at risk of social exclusion) in more than 70 schools distributed over ten states. Half of the children were selected randomly to participate in the program, with the remaining 50% assigned to the control group for the duration of the study (once the study finished, these children also joined the program). The program got underway at the beginning of the 2004-2005 school year and the outcomes for participants and the control group were evaluated at the end of the school year and later, at the beginning of the 2005-2006 school year.

In the short-term, the program showed a positive impact on academic performance and attitudes (dedication to study and behaviors), in contrast to aspects beyond the school environment (risk factors, family relationships, friendships and self-esteem). A significant part of these positive impacts however, including those associated with academic performance, were short-lived once the mentoring relationship finalized.

Subsequent studies with the same sample group of students have allowed observers to clarify in greater detail some of the program's effectiveness moderating elements. For example, it appears clear that a closer and longer-lasting relationship between the mentor and the child increases the program's impact margins, while at the same time, proves to be especially effective in children with a certain level of relational skills.

For more information:

Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J., Jucovy, L. Z. (2007). Making a Difference In Schools. The Big Brothers Big Sisters School-Based Mentoring Impact Study. Public/Private Ventures.

Herrera, C., Grossman, J. B., Kauh, T. J., McMaken, J. (2011). "Mentoring In Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring: School-Based Mentoring". Child Development, 82(1), 346-361.

Grossman, J. B., Chan, C. S., Schwartz, S. E. O., Rhodes, J. E. (2012). "The Test of Time In School-Based Mentoring: The Role Of Relationship Duration and Re-Matching on Academic Outcomes". American Journal of Community Psychology, 49(1-2), 43-54.

Bayer, A., Grossman, J. B., DuBois, D. L. (2013). School-Based Mentoring Programs: Using Volunteers to Improve the Academic Outcomes of Underserved Students. MDRC.

Schwartz, S. E. O., Rhodes, J. E., Chan, C. S., Herrera, C. (2011). "The Impact of School-Based Mentoring on Youths with Different Relational Profiles". Developmental Psychology, 47(2), 450-462.

Table 3 describes the characteristics of meta-analyses of the mentoring program effectiveness dealt with in this review. Three of these reviews focus on school-based mentoring programs; the remainder are included here as either a specific category of the program, or incorporated into the entire set (mixed) of programs reviewed.

Table 3.
School-based mentoring programs. Meta-analyses reviewed

Meta-analysis (reference country)	No. of studies included	Type of mentoring considered	Profile of children in programs	Profile of mentors	Dosage of mentoring	Effect measurement*
Bernstein *et al.* (2009) [33] (United States)	32	School-based	Primary & secondary students	Mixed profiles	Sessions per month (average) = 4.4 Average session duration= 1.1 hours. Average relationship extension = 5.8 months	-0.01 (prosocial behavior) -0.05 (maths) -0.04 (reading)
Dubois *et al.* (2002) [34] (United States)	55	School-based & community	Primary & secondary students	Mixed profiles	Relationship of +/- two hours per week; duration of +/- twelve months	+0.18 (global) +0.11 (academic)
Dubois *et al.* (2011) [35] (United States)	73	School-based & community	Primary & secondary students	Mixed profiles	From one twenty minute session per week for four months up to four two-hour sessions per week during an entire school year	+0.21 (global) +0.21 (academic)
Eby *et al.* (2008) [28] (United States)	112	School-based, community & occupational	Primary, secondary & university students	Mixed profiles	Varied dosages (not specified)	+0.36 (attitudes towards school) +0.19 (academic performance)
Wheeler *et al.* (2010) [36] (United States)	3	School-based	Primary & secondary students	Mixed profiles	Relationship starting with one hour per week with varied duration	+0.11 (learning behavior) -0.02 (maths) -0.01 (reading)
Wood & Mayo-Wilson (2012) [37] (United States)	6	School-based	Primary & secondary students	Mixed profiles	Between one-six hours per week and lasting between two and twelve months	+0.06 (attitudes) +0.09 (self-esteem) -0.01 (academic performance)

Source: Prepared by authors from https://educationendowmentfoundation.org.uk/uploads/pdf/Mentoring_Toolkit_references.pdf
* The standardized effect value is given, in accordance with Cohen's effect size measure (1988) [24]. In this way the measure of impact must be compared between programs. Based on Cohen's indications, the following is generally true: values similar to or less than 0.2 indicate a small effect size; values similar to 0.5, a medium effect size; values in the region of or greater than 0.8, a large effect size.

What does the accumulated evidence tell us about the effectiveness of school-based mentoring programs?

In general, it appears that the capacity of these programs to have a significant impact on students' academic outcomes tends to be quite limited. According to the summary of evidence from the Education Endowment Foundation, in general, the effect of these programs is similar to accelerating student academic course learning by, on average, one month, compared to mean academic gain of a school year. Moreover, when programs manage to generate significant positive impact (whether in terms of academic outcomes or in behavioral and attitudinal variables), the impact tends to fade a few short months after the mentoring relationship has finalized.

However, the effects of school-based mentoring can be highly variable, depending on the objectives taken into consideration, the characteristics of each program and the profile of the children mentored. More precisely:

- **Skills covered.** In general, school-based mentoring programs tend to be less effective when it comes to addressing non-cognitive attitudes and skills rather than strictly academic outcomes [35][38].
- **Dosage of mentoring relationship.** Several studies show the existence of a positive association between the duration of the mentoring relationship and its effectiveness [39][29]. For example, the BBBSA program appears to experience an increase in impact capacity when the mentoring bond lasts longer than nine months [40]. At the same time, the evidence shows that an unforeseen interruption in the mentoring relationship may have a negative impact on the child's academic and attitudinal outcomes [38], regardless of the initial planned duration of the relationship and even when the child has been reassigned to another mentor [39]. The literature is less conclusive regarding the potential association between the effectiveness of mentoring and the frequency of the mentor-child sessions. Some studies show that the probabilities of a program's success increase when mentor and child have at least one session per week [41].
- **Type of pairing.** Mentoring relationships appear to work better when the mentor and the child demonstrate a clear shared interest and when a close personal relationship is established between them [35][41]. On the other hand, the literature does not appear to be unanimous in relation to the added value of pairing mentors and children from the same ethnic group or of the same sex [33][34][35].
- **Structure of content.** In general, the evidence shows that mentoring relationships based on programs with well-structured activities and processes tend to function better than less structured mentoring interventions [34][42]. At the same time, some studies demonstrate that the impact capacity of school-based mentoring is diminished when the relationship between the mentor and the child is constrained due to an excessively restrictive task-focused planning [35].
- **Mentor profile.** The effectiveness of mentoring is increased when the occupational and training profile of the mentor corresponds to the contents and specific objectives the program is designed to address. When this is not the case, the specific training given to the mentor within the program framework may prove to be key [35].

- **Characteristics of treatment students.** Mentoring programs can be especially beneficial for young people who show behavioral problems and who come from a moderate, but not extreme level of social vulnerability [34][35][43]. In relation to gender and age, while social mentoring programs targeting young people tend to work better with older boys [35], school-mentoring appears to be more effective with younger girls [33][38].

School-based mentoring programs often tend to have little impact on students' academic outcomes and fade after a short period of time. These impacts can be quantified as accelerating academic process by one month compared to the mean gains for a school year.

Do peer tutoring programs work?

Peer tutoring forms a part of the teaching-learning process that takes place within the regular academic framework. These tutoring sessions are mainly conducted in pairs, with the tutor student being responsible for offering support to the treatment child and evaluating their learning process. Understood as a mechanism for addressing diversity, the objective of peer tutoring is generally two-fold: for the treatment child the aim is to improve specific cognitive skills (on occasion, also dealing with attitudinal skills); and for the student tutor, the aim is to work on certain metacognitive skills such as autonomy and responsibility.

These systems tend to differentiate between two similar major tutoring systems:

- *Peer tutoring between same age children.* In this system the tutor and tutee are students from the same school year, generally classmates. This category includes reciprocal peer tutoring programs, where students alternate roles of tutor and tutee, a procedure which shifts the logic of tutoring towards cooperative learning.
- *Cross-age tutoring.* Here, the role of the tutor is performed by students from higher course levels than those of the tutees. In this case, the type of relationship is basically asymmetric, and is often undertaken as a positive model strategy and compensatory supplement.

The duration and frequency of tutoring sessions can vary significantly between programs for both of the abovementioned systems. For example, samples with a higher dosage tend to schedule two or three tutoring sessions per week lasting approximately thirty minutes each over a period of between four months and the entire school year. In this way, peer tutoring programs can differ depending on the curriculum area covered (English, maths or others), as well as the social and academic profiles of tutors and tutees.

There are many individual tutoring programs which have been piloted experimentally, and below we explain the case of PALS (Peer-Assisted Learning Strategies) [44][45][46] program in the United States, and the Duolog [47] or Paired Reading [48][49] programs in the United Kingdom.

Box 3.
Peer-Assisted Learning Strategies (PALS) (United States)

Set up in 1989, PALS consists in programming a series of tutoring activities between classmates targeting gains in reading skills (PALS Reading) and maths (PALS Math). Through adopting a reciprocal system, PALS Reading is delivered in elementary, primary and secondary school classrooms, while PALS Math is delivered to primary school students. The purpose of PALS is to act as a supplementary resource to the students existing regular curriculum for these subjects.

With assistance from program materials and guidelines, the students supervise, correct, and encourage their partners. The tutor is awarded points in recognition of and depending on their partner's progress. The program contains several training resources for the teachers involved. PALS developers recommend a dosage of two or three 30 minute sessions weekly for reading, and from two to three sessions of similar duration for maths.

PALS is the peer-tutoring program that has generated the greatest volume of studies in the past twenty years, part of which are experimental. Forty classes in twelve primary and secondary schools participated in one of these experiments. Twenty of the classes were randomly assigned to the program –three tutoring sessions per week, lasting 35 minutes each over a 15-week period–, and the other twenty were assigned to the control group. After comparing outcomes in reading comprehension prior to and after participation in the program for both classes, the conclusions show that the intervention was moderately effective, and made an impact, regardless of students' initial skills level.

Experimental studies performed after the program have allowed researchers to corroborate the effectiveness of PALS Reading, also among special needs students; whereas other studies question the impact capacity of PALS Math.

For more information, see:
Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). "Peer-Assisted Learning Strategies: Making Classrooms More Responsive to Diversity". *American Educational Research Journal*, 34(1), 174-206.

Fuchs, L. S., Fuchs, D., Yazdian, L., Powell, S. (2002). "Enhancing First-Grade Children's Mathematical Development with Peer-Assisted Learning Strategies". *School Psychology Review*, 31(4), 569-583.

Sáenz, L. M., Fuchs, L. S., Fuchs, D. (2005). "Peer-Assisted Learning Strategies for English Language Learners with Disabilities". *ExceITPonal Children*, 71(3), 231-247.

There are two peer tutoring models: one between students of the same age (generally classmates) and another between students of different ages (when the tutor is from a higher level course than the tutee).

Table 4 provides the characteristics of meta-analyses on tutoring program effectiveness between similar programs dealt with in this article. Five of these meta-analyses include evaluations from both tutoring sessions between students of the same age as well as cross-age students. The remaining two focus their attention on tutoring systems between classmates.

Table 4.
Peer tutoring programs. Meta-analyses reviewed

Meta-analysis (reference country)	No. Studies included	Types of tutoring studied	Profile of student tutors	Profile of treatment students	Dosage of tutoring sessions	Effect measurement*
Bowman-Perrot *et al.* (2013) [50] (United States)	26	Classmates & cross-age	Not specified	Primary & secondary, mixed profiles	Overall time (average): 8 hours	+0.75 (global) +0.69 (primary) +0.74 (secondary)
Cohen *et al.* (1982) [51] (United States)	65	Classmates & cross-age	Primary & secondary; mixed profiles	Primary & secondary; mixed profiless	From one to thirty six weeks	+0.40 (over tutees) +0.33 (over tutors)
Cook *et al.* (1985) [52] (United States)	19	Classmates & cross-age	Primary & secondary; special needs	Primary & secondary; special needs	Average duration 9 weeks, 3 sessions per week and sessions lasting 24 minutes	+0.59 (for tutees) +0.65 (for tutors)
Ginsburg-Block *et al.* (2006) [53] (United States)	36	Classmates	Primary; mixed profiles	Primary; mixed profiles	Overall time (average): 15 hours	+0.35 (global) +0.48 (academic performance)
Jun *et al.* (2010) [21] (United States)	12	Classmates & cross-age	Secondary; mixed profiles	Secondary and adults; mixed profiles	Overall time: from 7 hours or less to 16 hours or more	+0.26 (global) +1.05 (cross-age) +0.92 (reading)
Leung (2015) [54] (United States)	72	Classmates & cross-age	Primary, secondary & university; mixed profiles	Primary, secondary & university; mixed profiles	Average duration 10 weeks, 2.5 sessions per week and 30 minute sessions	+0.39 (global) +0.50 (duration less than 10 weeks) +0.50 (duration more than 10 weeks)
Rohrbeck *et al.* (2003) [55] (United States)	90	Classmates	Primary; mixed profiles	Primary; mixed profiles	Average duration 15 weeks, 3.6 sessions per week and 45 minutes per session	+0.33 (global) +0.63 (same sex pairs)

Source: Prepared by Prepared by authors from https://educationendowmentfoundation.org.uk/uploads/pdf/Mentoring_Toolkit_references.pdf
* The standardized effect value is given, in accordance with Cohen's effect size measure (1988) [24]. In this way the measure of impact must be compared between programs. Based on Cohen's indications, the following is generally true: values similar to or less than 0.2 indicate a small effect size; values similar to 0.5, a medium effect size; values in the region of or greater than 0.8, a large effect size.

What can the evidence reviewed tell us about the impact capacity of peer tutoring programs?

It is clearly an established fact that peer tutoring can produce a significant positive impact on student learning outcomes. According to the summary of the Education Endowment Foundation, for students who participated in tutoring this impact was equivalent to gains of five months, on average, academic progress for students in a school year. The benefits of peer tutoring are also evidenced by academic outcomes and in the non-cognitive skills (attitudes, social and meta-regulation and emotional skills, etc.) of student tutors, both for peer tutoring between classmates as well as for cross-age peer tutoring [51][52][56]. The observation period for the studies reviewed was practically always confined to the same academic year in which the intervention took place and this is the reason why we cannot comment further on the impact capacity of these programs beyond that of the short-term.

The literature reviewed enables us to pinpoint exactly which characteristics of the programs and their participants act as modifiers of the effectiveness of the strategy we are dealing with here. The following points however, only make reference to the effectiveness of tutoring on the tutees.

- **Skills covered.** Peer tutoring (same or cross-age) have shown to effect a positive impact on the different curricular areas [54][55]. Nevertheless, a comparison between outcomes of the different studies would indicate that these interventions work better when it comes to tackling reading skills rather than maths or social sciences [46][50][57]. At the same time, these programs also demonstrate effectiveness when it comes to producing gains in specific non-cognitive outcomes (attitudes towards school, social and emotional skills, etc.) [53][58].
- **Dosage of program.** There is no evidence to show a clear relationship between the temporal tutoring system (including the total number of hours in the program) and its effectiveness [50] [55]. It is only in relation to the duration of a program that studies show that peer tutoring (same age or different ages) loses effectiveness when it is prolonged excessively over time (by way of guidance, when extended beyond 10-12 weeks) [51][54][58].
- **Type of pairing.** Evidence shows that, as is the case with other characteristics, tutoring systems where the student tutor is from a higher course than the tutee work better than tutoring between students of the same age [49][51]. Studies tend to assign greater impact capacity to tutoring programs between same age students involving reciprocal models [54]. Regardless of the tutoring system utilized, these tend to be particularly effective when students are paired with partners of the same gender [53][54][55].
- **Structure of contents.** Peer tutoring that has been designed using well-structured patterns of content and procedures work better than programs with little or no structure [51][53][54]. At the same time, it appears to be clear that these tutoring interventions lose effectiveness when utilized as a tool for introducing new content or content that is not closely associated with content that has been taught in regular class activity. By contrast, these work especially well as an instrument for supplementing or consolidating syllabus taught materials [51][58].

- **Mentor profile.** There is no evidence to support an evident association between the individual characteristics of the tutor and the impact capacity on students. Variables such as the tutor's academic level, their ethnic origin or socioeconomic status do not appear to condition, either in one way or another, the potential success of the programs [54]. On the other hand, it is not clear either whether tutor's training contributes any significant added value to tutoring impact capacity [51][54][58].
- **Characteristics of students mentored.** Despite the fact that globally students can benefit from peer tutoring, it is the most socially and academically vulnerable students who appear to gain most from intervention [51][53][58], together with students who have some form of disability [50][52][59] and younger students (primary school age) [51][53][49].

Peer tutoring can produce gains in academic outcomes equivalent to accelerating progress by five months compared to mean academic progress in a school year. The benefits can also be seen in the attitudes of tutees.

Summary

In general, the evidence reviewed indicates that ITPs can benefit students' academic progress. To the extent that they learn to address other disadvantaged students in an academic setting (and often socially), we could speak of potentially effective programs as a strategy for addressing diversity and equality of opportunities. At the same time, to a greater or lesser extent, the different individual tutoring systems demonstrate a positive impact on variables which are not strictly academic, such as motivation and attitudes towards schooling.

Even so, not all ITPs are equally effective, nor do they work in the same way for all student groups. The ranking from greater to lesser impact capacity would be: firstly, peer tutoring programs; secondly, tutoring intervention; and in third position, school-based mentoring. Peer tutoring programs are less costly that the rest of the ITPs (less costly than ITP schemes based on volunteering), at the same time as having the capacity to also effect a positive impact on the academic outcomes of the student tutors. This fact makes peer tutoring an especially cost-effective intervention model, with a cost-benefit balance that exceeds the rest of the one-to-one schemes [26][27][60].

This however does not question the potential of individual intervention programs or school-based mentoring to improve the academic perspectives of pre-schoolers and youth. The key lies in the capacity of these programs to align their designs (tutor profiles, contents and structuring of activities, dosage, etc.) with the needs of the intervention target groups.

The ranking from greater to lesser impact capacity would be: firstly, peer tutoring programs; secondly, tutoring intervention; and in third position, school-based mentoring.

Table 5 summarizes the advantages and limitations of the different intervention strategies reviewed.

Table 5.
Arguments for and against the ITPs reviewed

For	Against
Tutoring intervention	
Tutoring intervention can have a significant impact on students' performance	Evidence of the impact of these programs beyond the short-term is lacking
Evidence of effectiveness of these programs is solid in curricular material relevant to the area of reading and promising in maths	Evidence of the relevant curriculum area beyond those of language and maths is lacking
The most effective programs feature specialist teachers and are of a relatively intensive dosage	The cost of the most effective programs may be high (high volume of hours delivered by specialist teachers)
Tutoring delivered by volunteers can be effective when these individuals receive quality preparation and supervision	Training and supervision activities for voluntary tutors may represent significant administrative cost
Programs gain effectiveness when structured in line with schooling	Programs developed outside school hours can regular encounter difficulties in coordinating sessions with regular class activities and reference teachers
School-based mentoring	
School mentoring can make significant impact on non-cognitive skills (social, emotional, motivational)	The impact on academic outcomes tend to be modest and fade beyond the short-term
The most effective programs feature mentors with profiles interests well-aligned with the objectives of the relationship	Volunteering schemes, predominant in these programs, often tend to impede recruiting the most-suitable mentors
Mentoring carried out by volunteers can be effective when these individuals receive quality preparation and supervision	Training and supervision activities of volunteer mentors may incur significant administrative costs
The mentoring relationship gains effectiveness when it lasts for a minimally extensive time period and when a close bond is established between mentor and student	School-based mentoring tend to be constrained by the school calendar, a fact that may hinder temporary more effective schemes. Terminating a relationship before time has a negative effect on the mentored child's performance and attitudes
School-based mentoring is especially effective between younger girls and young boys with behavioral issues	School-based mentoring is not very effective in older girls with a profile of greater social vulnerability
Peer tutoring	
Peer tutoring shows relevant positive impact in performance and in non-cognitive skills (social, emotional of treatment students. these gains are also produced in the student tutor	Evidence of the impact of these programs beyond the short-term is lacking motivational)
Programs with tutors from higher level courses are especially effective, as well as reciprocal systems between classmates	Programs that pair students of different ages may generate difficulties in coordination
The benefits of peer tutoring are especially relevant in the area of reading	Peer tutoring appears to have less effect when applied in the area of maths
Peer tutoring works better when the design includes regular content and are used as an intervention mechanism	Peer tutoring is often used as a temporary and extraordinary resource, above all in schemes in which tutors are students from higher level courses
It appears that peer tutoring is especially effective in primary education, between students socially disadvantaged and with greater learning difficulties	Effectiveness of peer tutoring seems to be reduced in secondary education

Source: Prepared by authors.

Implications for practical application

As we already mentioned at the beginning of this article, the implementation of ITPs within the Catalan education system has been very disproportionate. We can state that peer tutoring is becoming a more and more frequent process in many schools, especially in pre-school and primary education. By contrast, tutoring intervention programs or school-based mentoring still remains relatively unchartered territory; and this despite the experience Catalonia has in developing other "beyond the school" initiatives (city-wide or district-wide educational projects, environmental schooling plans, second chance programs, learning communities, etc.).

The evidence tells us that ITPs can produce a positive impact on students' educational progress. It also tells us that, of the three ITP categories reviewed, peer tutoring demonstrates the greatest impact capacity and presents the most favorable cost-effectiveness and cost-benefit relationship. Finally, the evidence indicates that peer tutoring and school-based mentoring programs could work if the procedures and activity contents are aligned with schooling requirements and, eventually, behavioral needs targeted. Based on these confirmations, the best course of action appears to be:

- Reinforce and extend peer-tutoring experiences to include pre-school and primary, prioritizing: a) cross-age tutoring and reciprocal tutoring between students who are classmates; b) programs of short duration; c) programs addressing the area of reading comprehension.
- Make a commitment to intervention and school-based mentoring programs that have been subject to thorough diagnostic needs-analysis and which obey a well-founded theory of change. In the case of tutoring intervention, it would be recommendable to prioritize: a) programs based on well-structured curricula and closely associated with content covered in regular classwork; b) schemes where the role of the tutor is assigned to specialist teachers or volunteers who are suitably qualified in the subjects in question; c) intensive dosage patterns which last for a short or medium-term. In the case of school-based mentoring: establishing close mentor-student ties, which are significant and long-lasting; initiatives designed to address pre-school children and youth with academic or behavioral issues; programs that combine tutoring intervention with attitudinal work.

It would appear to be advisable to: 1) extend peer-tutoring experience to pre-school and primary; 2) commit to tutoring intervention and school-based mentoring programs based on a well-founded theory of change; 3) evaluate existing and potential ITPs.

Finally, it is necessary to evaluate the programs and initiatives implemented in the area of one-to-one tutoring, as these deal with evidence-based interventions or innovative initiatives with a less empirical foundation. It would therefore be necessary to promote the design and implementation of evaluations that enable us to comprehend how these programs work in practice, how effective they are globally and which components (contents, methodologies, dosage, the figure of the tutor, etc.), work best and for which student populations. Moreover, we must draw on this knowledge when it comes to fine-tuning the design of these programs or re-introducing them, thereby broadening the potential of their success.

Bibliography

[1] May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Tognatta, N. (2014). *Evaluation of the i3 Scale-up of Reading Recovery* (No. RR-79). CPRE. http://lfws.literacy.org/sites/default/files/researchreport/1488_readingrecoveryreport.pdf

[2] What Works Clearinghouse. (2013). *Reading Recovery*. Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_readrecovery_071613.pdf

[3] Ehri, L. C., Dreyer, L. G., Flugman, B., Gross, A. (2007). "Reading Rescue: An Effective Tutoring Intervention Model for Language-Minority Students Who Are Struggling Readers in First Grade". *American Educational Research Journal, 44*(2), 414-448.

[4] Gattis, M. N., Morrow-Howell, N., McCrary, S., Lee, M., Jonson-Reid, M., McCoy, H., Invernizzi, M. (2010). "Examining the Effects of New York Experience Corps® Program on Young Readers". *Literacy Research and Instruction, 49*(4), 299-314.

[5] Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. (2012). "The Effect of the Experience Corps® Program on Student Reading Outcomes". *Education and Urban Society, 44*(1), 97-118.

[6] Jacob, R., Armstrong, C., Willard, J. (2015). "Mobilizing Volunteer Tutors to Improve Student Literacy: Implementation, Impacts, and Costs of the Reading Partners Program". MDRC. http://www.mdrc.org/sites/default/files/Mobilizing_Volunteer_Tutors_FR.pdf

[7] Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., O'Connor, R. E. (1997). "The Effectiveness of One-to-One Tutoring by Community Tutors for At-Risk Beginning Readers". *Learning Disability Quarterly, 20*(2), 126-139.

[8] Vadasy, P. F., Sanders, E. A., Peyton, J. A. (2006). "Code-Oriented Instruction for Kindergarten Students at Risk for Reading Difficulties: A Randomized Field Trial with Paraeducator Implementers". *Journal of Educational Psychology, 98*(3), 508-528.

[9] Miller, S., Connolly, P. (2012). "A Randomized Controlled Trial Evaluation of Time to Read, a Volunteer Tutoring Program for 8- to 9-Year-Olds". *Educational Evaluation and Policy Analysis, 35*(1), 23-37.

[10] Miller, S., Connolly, P., Maguire, L. K. (2012). "The Effects Of a Volunteer Mentoring Programme on Reading Outcomes Among Eight-To Nine-Year-Old Children: A Follow Up Randomized Controlled Trial". *Journal of Early Childhood Research, 10*(2), 134-144.

[11] Maxwell, B., Connolly, P., Demack, S., O'Hare, L., Stevens, A., Clague, L. (2014). *TextNow Transition Programme. Evaluation Report and Executive Summary*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/uploads/pdf/EEF_Evaluation_Report_-_TextNow_Transition_-_October_2014.pdf

[12] Gorard, S., See, B. H., Siddiqui, N. (2014). Switch-on Reading. *Evaluation Report and Executive Summary*. Education Endowment Foundation. http://educationendowmentfoundation.org.uk/uploads/pdf/FINAL_EEF_Evaluation_Report_-_Switch-on_-_February_2014.pdf

[13] Rutt, S. (2014). *Catch Up Literacy. Evaluation Report and Executive Summary*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/uploads/pdf/Catch_Up_Literacy_(Final).pdf

[14] Rutt, S., Easton, C., Stacey, O. (2014). *Catch Up Numeracy- Evaluation Report and Executive Summary*. Education Endowment Foundation. http://www.nfer.ac.uk/publications/EFCU01/EFCU01.pdf

[15] Torgerson, C., Wiggins, A., Torgerson, D., Ainsworth, H., Hewitt, C., (2012). "The Effectiveness of an Intensive Individual Tutoring Programme (Numbers Count) Delivered Individually or to Small Groups of Children: A Randomised Controlled Trial". *Effective Education, 4*(1), 73-86.

[16] Torgerson, C., Wiggins, A., Torgerson, D., Ainsworth, H., Hewitt, C. (2013). "Every Child Counts: Testing Policy Effectiveness Using a Randomised Controlled Trial, Designed, Conducted and Reported to Consort Standards". *Research in Mathematics Education, 15*(2), 141-153.

[17] Ritter, G., Maynard, R. (2008). "Using the Right Design To Get The 'wrong' Answer? Results of a Random Assignment Evaluation of a Volunteer Tutoring Programme". *Journal of Children's Services, 3*(2), 4-16.

[18] Cook, P. J., Dodge, K., Farkas, G., Fryer Jr, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., Steinberg, L. (2014). "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago". NBER Working Paper No. 19862.

[19] D'Agostino, J. V., Murphy, J. A. (2004). "A Meta-Analysis of Reading Recovery in United States Schools". *Educational Evaluation and Policy Analysis, 26*(1), 23-28.

[20] Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How Effective Are One-To-One Tutoring Programs in Reading for Elementary Students at Risk for Reading Failure? A Meta-Analysis of the Intervention Research. *Journal of Educational Psychology, 92*(4), 605-619.

[21] Jun, S. W., Ramirez, G., Cumming, A. (2010). "Tutoring Adolescents in Literacy: A Meta-analysis". *Journal of Education, 45*(2), 219-238.

[22] Ritter, G. W., Barnett, J. H., Denny, G. S., Albin, G. R. (2009). "The Effectiveness of Volunteer Tutoring Programs for Elementary and Middle School Students: A Meta-Analysis". *Review of Educational Research, 79*(1), 3-38.

[23] Slavin, R. E., Lake, C., Davis, S., Madden, N. A. (2011). "Effective Programs for Struggling Readers: A Best-Evidence Synthesis". *Educational Research Review, 6*(1), 1-26.

[24] Cohen, J. (1988). *Statistical Power Analysis for The Behavioral Sciences*. London: Academic press.

[25] Wasik, B. A. (1998). "Using Volunteers as Reading Tutors: Guidelines for Successful Practices". *The Reading Teacher, 51*(7), 562-570.

[26] Washington State Institute for Public Policy. (2014a). *Tutoring: By Adults, One-on-One, Non-Structured*. Washington State Institute for Public Policy. http://www.wsipp.wa.gov/BenefitCost/ProgramPdf/370/Tutoring-By-adults-one-on-one-non-structured

[27] Washington State Institute for Public Policy. (2014b). *Tutoring: By Adults, One-on-One, Structured*. Washington State Institute for Public Policy. http://www.wsipp.wa.gov/BenefitCost/ProgramPdf/370/Tutoring-By-adults-one-on-one-structured

[28] Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., DuBois, D. L. (2008). "Does Mentoring Matter? A Multidisciplinary Meta-Analysis Comparing Mentored and Non-Mentored Individuals". *Mentoring, 72*(2), 254-267.

[29] Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J., Jucovy, L. Z. (2007). *Making a Difference in Schools. The Big Brothers Big Sisters School-Based Mentoring Impact Study*. Public/Private Ventures. https://bigsister.cmsservice.getfused.com/bigsister/file/Making%20a%20Difference%20in%20Schools.pdf

[30] Herrera, C., Grossman, J. B., Kauh, T. J., McMaken, J. (2011). "Mentoring in Schools: An Impact Study of Big Brothers Big Sisters School-Based Mentoring: School-Based Mentoring". *Child Development, 82*(1), 346-361.

[31] Rodríguez-Planas, N. (2012). "Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States". *American Economic Journal: Applied Economics, 4*(4), 121-139.

[32] Karcher, M. J. (2008). "The Study of Mentoring in the Learning Environment (SMILE): A Randomized Evaluation of the Effectiveness of School-based Mentoring". *Prevention Science, 9*(2), 99-113.

[33] Bernstein, L., Rappaport, C. D., Olsho, L., Hunt, D., Levin, M. (2009). *Impact Evaluation of the U.S. Department of Education's Student Mentoring Program. Final Report*. US Department of Education, NCEE Report 2009-4047. http://www.educationandemployers.org/wp-content/uploads/2014/06/impact-evaluation-us-mentoring-programme-ies.pdf

[34] DuBois, D. L., Holloway, B. E., Valentine, J. C., Cooper, H. (2002). "Effectiveness of Mentoring Programs For Youth: A Meta-Analytic Review". *American journal of community psychology, 30*(2), 157–197.

[35] DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., Valentine, J. C. (2011). "How Effective Are Mentoring Programs for Youth? A Systematic Assessment of the Evidence". *Psychological Science in the Public Interest, 12*(2), 57-91.

[36] Wheeler, M. E., Keller, T. E., DuBois, D. L. (2010). "Review of Three Recent Randomized Trials of School-Based Mentoring: Making Sense of Mixed Findings". *Social Policy Report. Volume 24, Number 3*. Society for Research in Child Development. http://eric.ed.gov/?id=ED519242

[37] Wood, S., Mayo-Wilson, E. (2012). "School-Based Mentoring for Adolescents: A Systematic Review and Meta-Analysis". *Research on Social Work Practice, 22*(3), 257-269.

[38] Rodríguez-Planas, N. (2014). *Do Youth Mentoring Programs Change the Perspectives and Improve the Life Opportunities of At-Risk Youth?* IZA World of Labor. Recuperat a partir de http://wol.iza.org/articles/do-youth-mentoring-programs-change-the-perspectives-and-improve-the-life-opportunities-of-at-risk-youth

[39] Grossman, J. B., Chan, C. S., Schwartz, S. E. O., Rhodes, J. E. (2012). "The Test of Time in School-Based Mentoring: The Role of Relationship Duration and Re-Matching on Academic Outcomes". *American Journal of Community Psychology, 49*(1-2), 43-54.

[40] Herrera, C. (2004). *School-Based Mentoring. A Closer Look*. Public/Private Ventures. http://www.vamentoring.org/images/uploads/resources/PPV_School_Based_Mentoring_A_Closer_Look_2004.pdf

[41] Bayer, A., Grossman, J. B., DuBois, D. L. (2013). *School-Based Mentoring Programs: Using Volunteers to Improve the Academic Outcomes of Underserved Students*. MDRC. http://eric.ed.gov/?id=ED545462

[42] Langhout, R. D., Rhodes, J. E., Osborne, L. N. (2004). "An Exploratory Study of Youth Mentoring in an Urban Context: Adolescents' Perceptions of Relationship Styles". *Journal of Youth and Adolescence, 33*(4), 293-306.

[43] Schwartz, S. E. O., Rhodes, J. E., Chan, C. S., Herrera, C. (2011). "The Impact of School-Based Mentoring on Youths with Different Relational Profiles". *Developmental Psychology, 47*(2), 450-462.

[44] Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). "Peer-Assisted Learning Strategies: Making Classrooms More Responsive to Diversity". *American Educational Research Journal, 34*(1), 174-206.

[45] Fuchs, L. S., Fuchs, D., Yazdian, L., Powell, S. (2002). "Enhancing First-Grade Children's Mathematical Development with Peer-Assisted Learning Strategies". *School Psychology Review, 31*(4), 569-583.

[46] What Works Clearinghouse (2012). *Peer-Assisted Learning Strategies.* Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_pals_060512.pdf

[47] Topping, K. J., Miller, D., Murray, P., Henderson, S., Fortuna, C., Conlin, N. (2011). "Outcomes in a Randomised Controlled Trial of Mathematics Tutoring". *Educational Research, 53*(1), 51-63.

[48] Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., Ahmed, H. (2015). *Paired Reading. Evaluation report and Executive summary.* Education Endowment Foundation. https://educationendowmentfoundation.org.uk/uploads/pdf/Paired_Reading.pdf

[49] Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., Miller, D. (2011). "Improving Attainment across a Whole District: School Reform through Peer Tutoring in a Randomized Controlled Trial". *School Effectiveness and School Improvement, 22*(3), 265-289.

[50] Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., Parker, R. (2013). "Academic Benefits of Peer Tutoring: A Meta-Analytic Review of Single-Case Research". *School Psychology Review, 42*(1), 39–55.

[51] Cohen, P. A., Kulik, J. A., Kulik, C.-L. C. (1982). "Educational Outcomes of Tutoring: A Meta-analysis of Findings". *American Educational Research Journal, 19*(2), 237-248.

[52] Cook, S. B., Scruggs, T. E., Mastropieri, M. A., Casto, G. C. (1985). "Handicapped Students as Tutors". The Journal of Special Education, 19(4), 483-492.

[53] Ginsburg-Block, M. D., Rohrbeck, C. A., Fantuzzo, J. W. (2006). "A Meta-analytic Review of Social, Self-Concept, and Behavioral Outcomes of Peer-Assisted Learning". *Journal of Educational Psychology, 98*(4), 732-749.

[54] Leung, K. C. (2015). "Preliminary Empirical Model of Crucial Determinants of Best Practice for Peer Tutoring on Academic Achievement". *Journal of Educational Psychology, 107*(2), 558-579.

[55] Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., Miller, T. R. (2003). "Peer-assisted Learning Interventions with Elementary School Students: A Meta-Analytic Review". *Journal of Educational Psychology, 95*(2), 240-257.

[56] Britz, M. W., Dixon, J., McLaughlin, T. F. (1989). "The Effects of Peer Tutoring on Mathematics Performance: A Recent Review". *Journal of Special Education, 13*, 17-33.

[57] Lloyd, C., Edovald, T., Morris, S., Kiss, Z., Skipp, A., Haywood, S. (2015). *Durham Shared Maths Project. Evaluation report and Executive summary.* Education Endowment Foundation. https://educationendowmentfoundation.org.uk/uploads/pdf/Shared_Maths_1.pdf

[58] Robinson, D., Schofield, J., Steers-Wentzell, K. (2005). "Peer and Cross-Age Tutoring in Math: Outcomes and Their Design Implications". *Educational Psychology Review, 17*(4), 327-362.

[59] Sáenz, L. M., Fuchs, L. S., Fuchs, D. (2005). "Peer-Assisted Learning Strategies for English Language Learners with Disabilities". *Exceptional Children, 71*(3), 231-247.

[60] Washington State Institute for Public Policy (2014c). *Tutoring: By Peers.* Washington State Institute for Public Policy. http://www.wsipp.wa.gov/BenefitCost/ProgramPdf/107/Tutoring-By-peers

What grouping strategies respond to criteria of efficiency and equality?

Gerard Ferrer-Esteban

Researcher at the Giovanni Agnelli Foundation and the Interdisciplinary Group on Education Policies (GIPE) at the University of Barcelona

Motivation

In international circles, several school student grouping strategies and practices are used to address diversity and respond to the difficulties associated therein. There appears to be a consensus with regard to the fact that a reduction in the student-teacher ratio, an implicit feature of grouping strategies, facilitates the task of instruction and addressing diversity. Even so, there are several different approaches as to how and according to what criteria students should be grouped. Each approach brings with it a series of implications in terms of school organization and can generate different outcomes in terms of learning.

Diversity can be managed, for example, by way of "differentiation methods", which include ability tracking or streaming between classes and intra-classroom ability grouping or setting, which in the Catalan case is similar to methods we refer to as "flexible grouping". The aim of these strategies is to create homogenous spaces to facilitate the process of instruction through objective-focused teaching and content that is better aligned with students' abilities.

On the other hand, centers also group students in line with heterogeneous criteria, principally according to their skills, abilities and other observable characteristics. In these cases, diversity is often dependent on the organizational conditions and functioning of the learning groups. Strategies such as cooperative learning are founded on the internal heterogeneous nature of groups to carry out strategies that benefit from students' diversity of profiles and skills (positive inter-dependence, students mutually assisting one another, peer effects, etc.).

There are different approaches as to how students should be grouped and what criteria should be used. Each approach brings with it different implications in terms of school organization and produces varying learning outcomes.

Despite resistance from a large section of teachers and families, grouping by levels has become an ever more popular strategy being used in primary & secondary school education centers in Catalonia to meet the growing demand of student diversity. In the same way, even with greater level of acceptance, cooperative learning has consolidated itself as a strategy for responding to the difficulties involved in managing the diversity facing teachers in the context of class-groups. However, many questions concerning each strategy remain unanswered and which can be addressed by an evidence-based analysis.

Box 1.
Type of grouping – key concepts

Grouping strategies. Practices which are made up of work groups smaller than those of regular sized class groups. This kind of grouping may be of a more or less punctual nature, academically homogenous or heterogeneous and can apply cooperative educational activities or of another variety.

Ability tracking or streaming. Ability tracking or streaming. Stable class groups which are made up of students with the same academic level. This method is similar to within-class grouping by levels implemented in some Catalan secondary schools. Usually the lower level groupings are of a more reduced class size than the higher levels.

Intra-classroom ability grouping or setting. Reduced academically homogenous groups made up of students from the same reference class group (or different class groups from the same course) to work on specific materials. They are not permanent by nature and the students can transfer from one group to another. They are comparable to the flexible groups organized by a good number of primary and secondary schools in Catalonia to address linguistic and mathematical issues.

Multi-graded and multi-age grouping (non-graded or ungraded program, cross-age or cross-grade grouping). Reduced groups or class groups, same academic levels and made up of students of different ages, normally from the same academic cycle (3 year bands). These tend not to be permanent and are organized to work on one or several specific subjects. In Catalonia, this strategy is only implemented in a few education centers.

Heterogeneous grouping. Reduced groups with different levels made up of students from the same reference class group, of a flexible nature and not permanent, designed to work on specific subjects. These groups can apply strategies of cooperative work (positive inter-dependence, mutual assistance, common learning objectives, etc.) or other types of modalities.

Questions influencing the review

Grouping came about as a strategy designed to break with the more traditional education model based on instruction delivered to the entire class group. In order to address this issue, we have divided the questions influencing the review into four separate blocks. The first two allow us to introduce the debate about grouping, and the second two will deal with specific grouping modalities:

- The first question we will respond to is whether reducing class size, and in turn, the **student-teacher ratio**, has any significant effect on students' learning. In the event that there is a significant effect, does this effect compensate for the high cost involved?
- In second place, prior to undertaking an analysis of any specific grouping strategies, we will consider the debate about whether **to group or not to group students**: is working in small groups an effective strategy with regard to teaching models implemented with the entire class? Are there any conditions under which grouping strategies are more effective than traditional models?
- Thirdly, we will look at the strategies dealing with **ability tracking or streaming**: can we be sure that they produce cognitive gains for all students? In the event that the answer is yes, are the expected benefits distributed equally among students in groups with different learning objectives and content? Regarding students with lower performance levels, who are often characterized as having low learning expectations, poor motivation and low self-esteem when it comes to learning, to what extent do these students benefit from being grouped with other students of a similar profile?
- Finally, the review focuses on strategies implemented within **heterogeneous grouping**: are there effective strategies for assisting learning in complex settings where a huge diversity of learning skills, abilities and expectations overlap? If there are, how are these benefits distributed between students with different skills? Are the expected gains for students with greater learning difficulties in detriment of those for students with more ability?

Reviewing the evidence

For this review we have chosen a total of seventeen meta-analyses and three non-systematic reviews, covering over 500 studies in dealing with the effects of class size, student grouping and grouping strategies. In this review we will discuss, in the following order, class size, grouping versus non-grouping; ability tracking or streaming versus non-grouping and heterogeneous grouping; and grouping with cooperative strategies versus non-grouping or non-structured cooperative grouping.

A preview: the debate about class size

The debate about grouping strategies is preceded by another debate that has marked the educational policy agenda around the world over the past fifty years: class size.

The first meta-analyses, carried out during the 1970s and 1980, indicated that a reduction in class size was associated with improved teaching effectiveness and attitudes, referring to both teachers and students, which were better focused on learning, while at the same time pointing out the importance of reducing ratios in order to facilitate greater individualization of instruction [1]

Research shows that a reduction in class size can produce medium and long-term gains in terms of learning and educational continuity.

This brings us up to the present day and all the efforts that have been made to study the effect of a reduction in class size, one research program promoted in 1985 marks a turning point: the STAR (Student-Teacher Achievement Ratio) program implemented in Tennessee, United States. This experimental program was based on randomly assigning teachers and students into small class groups (13-17 students), larger class groups (22-26 students) or into large classes with a teacher's assistant. As a result of this program, a huge volume of research has been generated which consistently points to the fact that a reduction in class size can produce benefits in terms of learning and educational continuity in the medium and long-term [2] [3] [4].

The benefits derived from the reduction in class size are estimated approximately as being similar to gains of three months academic progress on average, in favor of the classes with fifteen students, compared with the classes with an average of twenty-two students [5] [6]. As we can see in Table 1, the positive sense of the learning effect is reflected in other studies carried out using quasi-experimental designs in other parts of the world, such as for example in Israel [7], France [8], Sweden[9] and Kenya [10].

The key is if the reduction in ratio is great enough to produce significant changes in students' learning styles, and consequently, in the outcomes.

Nevertheless, other quasi-experimental research has found that these effects may also produce little or insignificant effects [11] [12] [13] [14]. According to the summary of the Education Endowment Foundation the key is if the reduction is great enough to produce significant changes in the students' styles of learning, and consequently, in outcomes. Unless the reduction in ratio is very significant, having smaller sized classes does not necessarily mean that the teacher will modify teaching methods or make any significant changes in the use of class time[15]. It is therefore understood that any reduction that is not below twenty or even below fifteen students per teacher, if unlikely to translate into substantial gains in students cognitive and non-cognitive outcomes [16].

Table 1.
Evaluations of class-size reduction programs (N=8)

Study (reference country)	Research design	Level of education	Skills	Average (AE) & differential effects (DE)		Standard deviation
Nye & Hedges (2002) [3] (United States)	Experimental	Primary school	Language & maths	AE: Performance in students 0.19 who perform poorly	Signif.	0.16-0.19
				DE: Performance in students poorly* small classes	n/s	-
Nye *et al.* (2000) [5] (United States)	Experimental	Pre-school & primary	Language & maths	AE: Performance	Signif.	0.11-0.30
Nye *et al.* (1999) [4] (United States)	Experimental	Primary school 0.22 lower secondary	Language & maths		Signif.	0.13-0.22
Krueger & Whitmore (2001) [2] (United States)	Experimental / Heckman correction	Lower secondary education	University entrance exams	AE: Performance	Signif.	0.13
				DE: Afro-American	Signif.	0.20
Dulo *et al.* (2012) [10] (Kenya)	Quasi-experimental	Primary school	Language & maths	AE: Performance	Signif.	0.05-0.06
Bressoux *et al.* (2008) [8] (France)	Quasi-experimental	Primary school	Language & maths	AE: Performance	Signif.	0.03
				DE: High level students	Signif.	0.02-0.03
				DE: Low level students	Signif.	0.02-0.04
Hoxby (2000) [14] (United States)	Quasi-experimental	Primary school	Language & maths	AE: Performance	n/s	-
Angrist & Lavy (1999) [7] (Israel)	Quasi-experimental	Primary school grades 5 & 4 (language)	Language & maths	AE: Performance	Signif.	0.13-0.18
		Primary school grade 3 & 4 (maths)		AE: Performance	n/s	-

Source: Prepared by the authors based on Impact Evaluations in Education (World Bank).
n/s: no significant effect.

Research shows that a reduction in class size can produce medium and long-term gains in terms of learning and educational continuity.
The key is if the reduction in ratio is great enough so as to produce significant changes in students' learning styles, and consequently, in the outcomes.

Is grouping students a worthwhile alternative to traditional teaching?

Taking into account skepticism regarding the importance of the effects of reducing class size, one very plausible conclusion is that the cost of such policies can outweigh the benefits obtained. If the aim is to respond to criteria of effectiveness, but also efficiency, investment in policy to address diversity must surely have to promote other strategies which also focus on reducing the student-teacher ratio, but which are compatible with class groups of more than twenty students (and which are as such, economically more sustainable). Some of these strategies might entail individual tutoring programs, others still, student grouping programs, which are the initiatives we will look at in this review.

Before beginning an evaluation of grouping strategies, we will begin by comparing the average effect of grouping students with that of teaching an entire class group. In order to do so, we will use two recent and methodologically "solid" meta-analyses, taking into account strict inclusion criteria of the studies: experimental studies using random assignment or quasi-experimental designs with experimental groups and test-based monitoring prior to and after the intervention.

Grouping students within the class can produce a positive impact on learning, more than teaching the entire class group.

The evidence leaves no room for doubt regarding the effectiveness of grouping. The two meta-analyses indicate that grouping students within the class can produce a positive impact on learning, more than teaching the entire class group. This positive impact is evident both by observing the cognitive outcomes (academic performance in reading and maths) [17] [18] as well as non-cognitive skills (general self-concept and attitudes towards the subject) [18]. The effects on reading comprehension would be comparable to half a year's academic progress, taking into account the average performance growth observed in reading annually [17] [19]. All students, especially those with greater learning difficulties, benefit from grouping strategies [18].

Grouping for the sake of grouping, without applying differentiated instruction strategy with regard to teaching the entire class group, does not produce, on its own. any significant effect.

This represents an average effect behind which are hidden differential effects depending on didactic strategies or the quality of teaching staff (Table 2). For example, we can say that grouping for the sake of grouping, without applying any differentiated instruction strategy with regard to teaching the entire class group, does not produce, on its own. any significant effect [18]. The effects of grouping are more pronounced insofar as additional materials are used, teaching staff are better trained or strategies that reward students for their efforts and work are used [18] [20]. The results of the research also seem to confirm the idea that the positive impact of grouping is more significant in primary school and that, in order to see any effects, grouping must be implemented more than once per week [18].

The analyses "grouping vs. no grouping" also allows us to confirm that both homogenous grouping of a flexible nature [18] [21] as well as heterogeneous grouping [17] [18] seem to produce a positive impact on learning when they are compared, separately, with traditional instruction for the class group. So, what exactly are these impacts?

Table 2.
Grouping strategies. Meta-analyses reviewed

Meta-analysis (country of reference)	Grouping type	Number of studies included	Research design	Duration	Level of education	Skills considered	Average effect (AE) & differential effect (DE)*	
Puzio & Colby (2010) [17] (United States)	Heterogeneous within class vs. grouping	15	Experimental & quasi experimental	From 10-40 weeks (30-90 minutes)	Primary & secondary	Reading	AE: Performance	+0.22
Lou *et al.* (1996) [18] (United States)	Grouping within class vs. non- grouping	51	Experimental & quasi experimental	3 terms: < 4 weeks 4-16 weeks > 17 weeks	Primary & secondary & post-secondary	Reading maths & sciences	AE: Performance	+0.17
							DE: Low level of competency	+0.37
							DE: Medium level of competency	+0.19
							DE: High level of competency	+0.28
							DE: Maths class	+0.20
							DE: Language class	+0.13
							DE: Teacher's training	+0.42
							DE: Adapted materials	+0.26
							DE: Incentives	+0.29
							DE: Adapted instruction	+0.25
							DE: Cooperative learning	+0.28
							DE: Intensity (>1 session per week)	+0.22
							DE: 2nd cycle primary education	+0.29
		21	Experimental & quasiex-perimental	n/a	Primary & secondary & post-secondary	Attitude	AE: Towards subject	+0.18
							AE: towards instruction	-0.13 (n/s)
		10	Experimental & quasiex-perimental	n/a	Primary & secondary & post-secondary	Self-concept	AE: General	+0.16
							AE: Academic	(n/s)

Source: Prepared by authors based on Puzio and Colby (2010) & Lou et al. (1996)
* Standardized effect size (Cohen,1988). Small effect: 0.2. Medium effect: 0.5. Large effect: 0.8.

Ability tracking or streaming: effectiveness at the cost of fairness?

Ability tracking or streaming consists in separating students according to their level, either observed or potential, of ability and skills, to assign them into academically homogenous groups. Students can be grouped between classes (tracking or streaming), or within the same class group for specific subjects (ability grouping or setting). There is also the non-graded or ungraded program, cross-age or cross-grade grouping, in which students are grouped according to their abilities and regardless of age. The objective in grouping students with similar skills levels is to try to reinforce students' learning by better aligning the teaching objectives and content with students' learning abilities and needs. Some argue that this kind of grouping strategy should allow better and more efficient attention to diversity, given that it offers enhanced content and a higher learning pace for students with greater abilities and a more individualized attention to students with learning difficulties.

Some programs combine grouping by levels with other methodological approaches. This is the case of Reading Edge, a program developed in early secondary school education [22] both in the United States as well as in the United Kingdom [23] [24] [25], and which forms part of the Success for All school reform model. Reading Edge is a comprehensive reading program that allows teaching staff to accelerate the varying learning paces by using differentiated instruction, training feedback and ongoing evaluation [23]. Despite using flexible grouping, the basis for this program is cooperative learning. It was born out of the need to place greater emphasis on instruction and curricula content in the context of secondary level education, without abandoning the expected benefits of positive inter-dependence between students. Other programs implemented in the United States have contemplated the use of homogenous flexible grouping within the framework of different intervention models. Here we are referring to programs such as Horizons [26] and Distar Arithmetic [27] within the framework of the school reform model Direct Instruction, or Literacy Collaborative, based on self-instruction strategies and the use of heterogeneous and homogenous learning groups [28] [27].

Table 3.
Homogenous grouping strategies. Meta-analyses reviewed

Meta-analyses (country of reference)	Type of grouping	Number of studies included	Research design	Duration of program	Level of education	Skills considered	Average effects (AE) & differential effects (DE)**		Differential effects according to level of skills**		
									Low	Med.	High
Lou *et al.* (1996) [18] (United States)	Homogenous within the class vs. heterogeneous groups	12	Exp. & quasi-ex.*	n/a.	Primary, secondary or post-secondary	Reading, maths & sciences	AE: Performance	+0.12	-0.60	+0.51	+0.09
							ED: Maths class	0.00			
							ED: Language class	+0.36			
Kulik & Kulik (1992) [29] [30] (United States)	Homogenous between classes	56	Quasi-ex.	Between 1 semester & 3 years	Primary & Secondary	Reading, maths, sciences & social studies	AE: Performance	+0.03	-0.01	-0.02	+0.10
	Multi-grade	14	Quasi-ex.	Between 1 & 3 years	Primary	Reading	AE: Performance	+0.30	+0.29	-0.01	+0.12
	Homogenous within the class	11	Quasi-ex.	Between 6 weeks & 1 year	Primary & Secondary	Reading & maths	AE: Performance	+0.25	+0.16	+0.18	+0.30
Slavin (1990)*** (United States) [31]	Homogenous between classes or within the class	15	Exp. & quasi-ex.	Between 1 semester & 5 years	Secondary	Reading, maths, sciences & social studies	AE: Performance	-0.06	-0.06	-0.10	+0.05
							ED: Homogenous between classes (n=13)****	-0.03	-0.05	-0.08	+0.01
Gutiérrez & Slavin (1992) [32] (United States)	Multi-grade in one subject	7	Quasi-ex. (1 exp.)	Between 1& 3 years (1 study, 5 years)	Primary	Reading & maths	AE: Performance	+0.46			
	Multi-grade in all subjects	14	Quasi-ex.	Between 6 months &3 years (2: 5 & 6 years)	Primary	Reading & maths	AE: Performance	+0.34			
Slavin (1987) [21] (United States)	Homogenous between class	14	Quasi-ex. (2 exp.)	Between 1 semester & 4 years	Primary	Reading & maths	AE: Performance	-0.02			
	Multi-grade	12	Quasi-ex. (2 exp.)	Between 1& 3 years	Primary	Reading	AE: Performance	+0.45			
	Homogenous within class	5	Exp.	Between 4 & 8 months	Primary	Maths	AE: Performance	+0.32	+0.66	+0.28	+0.35

Source: Prepared by author based on Kulik & Kulik (1992), Slavin (1990), Gutiérrez & Slavin (1992) and Slavin (1987).
* Quasi-ex.: quasi-experimental. Exp.: experimental.
** Standardized effect size (Cohen,1988). Small effect: 0.2. Medium effect: 0.5. Large effect: 0.8.
*** This meta-analysis (Slavin, 1990) includes a total of 29 studies, of which 6 are experiments using random assignment, 9 studies using grouping techniques (matching) and 14 studies based on correlations. In this review of reviews only the effects identified for the first two types of studies were used.
**** Effect size measurements calculated only using studies which analyzed homogenous grouping between classes (Slavin, 1990: table 1).
n/a: Not available.

What does the evidence tell us about the effectiveness & equity of ability tracking or streaming?

Ability tracking or streaming has a practically no effect on academic performance, in primary or secondary education.

It is safe to say that **ability tracking or streaming has practically no effect on academic performance, in primary** [29] [21], **or secondary education** [29] [31]. While some studies conclude that there is no differential effect between students grouped in classes of greater or lesser academic ability [29] [31], a significant portion of evaluations have observed a tendency among these grouping strategies to benefit students in the higher skills groups and to undermine those in the lower skills ability groups (even though they continue to indicate that there is no significant overall effect) [33] [34] [35] [36] [37].

Homogenous grouping within classes favors medium levels of learning, however, in comparison with heterogeneous grouping, it tends to prejudice student with lower learning abilities.

As far as homogenous grouping within the class is concerned, this strategy seems to favor medium learning levels [18] [29] [30] [21].[1] This grouping modality is characterized by a high level of formal organizational flexibility that permits student mobility between heterogeneous class groups.

Insofar as equity is concerned, the evidence presents conflicting results depending on the groups being compared. If the comparison is made with a class group where no grouping is undertaken, the effects on the homogenous grouping within the class are positive [30] [21] [38]. On the other hand, if the comparison is made with a class group subject to heterogeneous grouping, the effect of the ability tracked groups tends to hinder students who are lower attaining learners [18]. As the Education Endowment Foundation indicates in the summary, students classified as lower attaining learners may lose out, one or two months each year, in terms of academic progress when compared with heterogeneous groups. On the other hand, students with higher academic ability produce gains amounting to two additional months' academic progress.

Based on the evidence, the conditions required by which it would be recommendable to implement ability tracking or streaming in order to favor learning and mitigate the risk of impairing more vulnerable students are the following:

- **Grouping modality.** All the accumulated evidence allows us to exclude any form of grouping between classes and highlights the effectiveness of homogenous grouping within classes or multi-age grouping (Table 3). These grouping models are more effective than instruction with class groups for students with greater difficulties [30] [21] [38], but prove less effective for this student profile when compared with heterogeneous grouping [18].

1 This effectiveness can also be observed in homogenous multi-grade or multi-age grouping strategies [32] [21] [30].

- **Curricular and instruction pace adjustments.** Programs that introduce more significant adjustments to the curriculum varying the pace of instruction in a coherent manner and adapting the materials to suit the skills of each group which produce positive effects [21] [30] [29] [31].
- **Grouping flexibility & student mobility.** A key difference between how tracking between classes and grouping strategies within classes or multi-grade grouping work is that the former are functionally extremely rigid and do not allow for student mobility between groups. Strategies which provide greater mobility between groups are more effective, although the more inclusive models produce greater benefits [32] [21] [39].
- **Ongoing evaluation.** In order to permit mobility between differentiated level groups, it is indispensable to have an ongoing evaluation system, as is the case in the Reading Edge program for example on a regular basis [23] [31].
- **Utilization of homogenous grouping for no more than one or two subjects.** Students should regularly be placed in the corresponding heterogeneous class group and be grouped by levels for subjects when deemed necessary (very often language and maths) [21].
- **Combination with other methodologic approaches.** Programs like Reading Edge combine differentiated instruction in homogenous reading groups (by level and multi-grade) with cooperative learning [40] [23] [24].

Cooperative learning in heterogeneous grouping: equity at the cost of effectiveness?

Grouping students facilitates the development of teaching and learning strategies that are impossible in class groups of twenty-five students. Several of these strategies, as is the case with cooperative or collaborative learning, allow the active participation of all students in the individual learning process, while at the same time encouraging interchange and cognitive conflict between students working together to achieve common learning goals [41] [42]. Students can work on a specific part of the assigned task to contribute to a common outcome, or alternatively work together in a shared activity throughout the entire session.

These strategies are often presented as an alternative to flexible grouping by levels, given that in the majority of cases, cooperative learning groups are put together based on the heterogeneous nature of the students. In fact, diversity very often becomes one of the conditions of the organizational and functional features of learning groups: based on internal heterogeneity of the groups so that the development of the activity can benefit from the students' profile diversity and skills. Effective collaboration depends on the extent to which positive inter-dependence and a relationship of mutual assistance exists between the members in the group [43].

There is great diversity of strategies in cooperative groups, whose development can vary depending on the degree of structuring (oriented versus common objectives), the materials used, teacher profiles (training, experience), the combination with other didactic strategies (direct instruction) or on the characteristics of the students (age, abilities towards learning).

Currently, there are several specific cooperative learning programs which have been evaluated, even though with different outcomes, by experimental and quasi-experimental designs. In Table 1 we illustrate and briefly explain the CIRC (Cooperative Integrated Reading and Composition) as well as its bilingual teaching variant, BCIRC (Bilingual Cooperative Integrated Reading and Composition), developed in the United States and in the United Kingdom starting at second grade primary school level up to fifth or sixth grade [44] [45] [46] [47] [48] [49]. In these two countries we can also find the STAD (Student Teams-Achievement Divisions) [62] [69] Team Assisted Individualization (TAI) programs, which combine individualized attention with collaborative group work [50] [69]. Other programs that have demonstrated their effectiveness in secondary education include Student Team Reading and Writing [51] [52] [53] and the Reading Edge program which, despite adopting a cooperative learning structure, organizes homogenous grouping according to the reading abilities of secondary school students.

In the area of primary school math, we find the TAI Math and STAD programs, which at secondary school level is referred to as PowerTeaching:Mathematics [54] [20]. At secondary level there is IMPROVE [55] [56], a program designed in Israel which combines cooperative learning with meta-cognitive instruction. Other specific cooperative learning programs include Core-Plus Mathematics [57], Connect Mathematics [58] and the Interactive Mathematics Program, a 4-year program for learning algebra through precalculus [59].

Catalonia has a long-standing pedagogic history associated with cooperative learning and there are several initiatives being developed at education centers. One of the most notable is the CA/AC program - *"Cooperar per Aprendre/Aprendre a Cooperar"* – (Cooperate to Learn/Learn to Cooperate), developed by the Research group for Diversity (GRAD, as per the Catalan acronym) at the University of Vic-Central University of Catalonia (UVic-UCC). The program is designed for pre-school, primary and secondary school teaching professionals, through the Educational Innovation & Training Centre (CIFE) working group at UVic-UCC about "Inclusive education, cooperation between students and collaboration between teachers", closely associated with the GRAD. The aim of this training and counselling program is to facilitate schools' access to the tools and resources required to modify teaching methods and the structure of class activities, as well as instructing students on how to work in a team [60]. Currently, there are more than 450 education centers throughout Spain applying these resources.

What can be said about the effectiveness of cooperative learning? Is there evidence to support the benefits in terms of learning? How are these benefits distributed?

The most recent meta-analyses confirm to a great degree the evidence already mentioned by the volume of studies carried out during the eighties (Graph 1): cooperative learning strategies have a positive impact on students' performance in basic math, reading & science skills [18] [61] [62] [63] [64] [20] [65]. According to the summary document prepared by the Education Endowment Foundation, the average gains produced by this type of strategies can equal up to five months in terms of academic progress. Apart from the

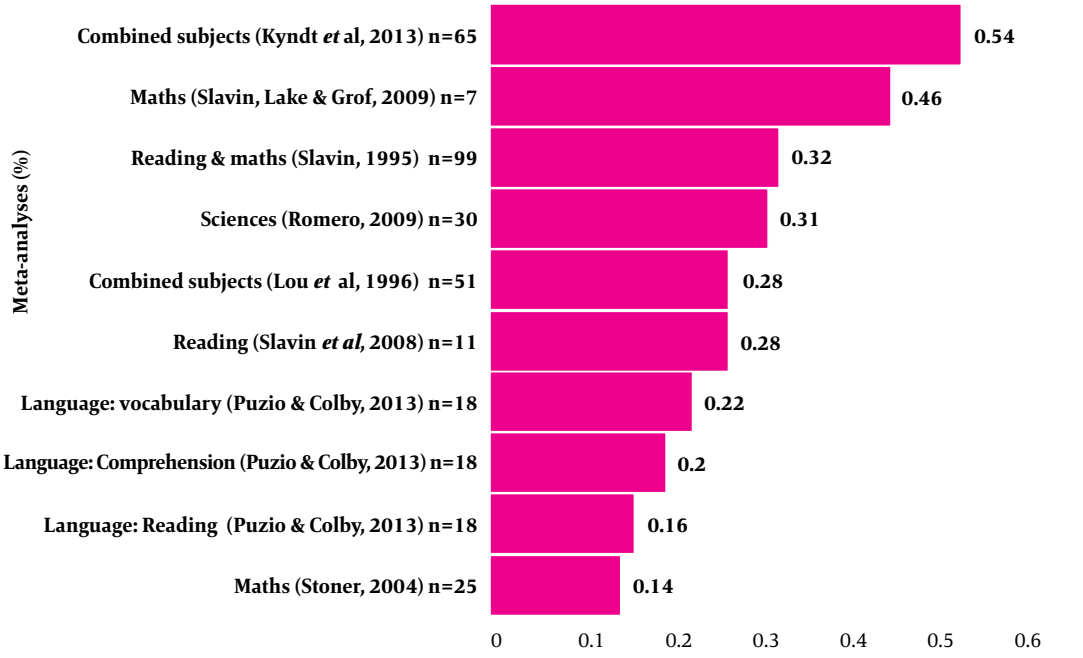
effects on academic performance, this impact can also be observed in non-cognitive outcomes, such as attitudes and perceptions towards learning and the subject in question [65].

Even so, not all cooperative learning strategies produce the same impact (Table 4 and Graph 2). Their effectiveness depends on the design and structure of the intervention, context or level of education in which the program is implemented. It is possible therefore, based on the evidence, to indicate several conditions:

- **Group objectives and individual accountability.** Effectiveness of cooperative activities increase when groups work towards achieving a common goal and obtain some form of recognition [20] [62]. Their actual success is tied to the success of others. At the same time, the success of the group must be dependent on the individual learning of each member and not on a product of the group [20].
- **Positive inter-dependence (functional).** The students who obtain greater benefits from cooperative groups are those who give and receive elaborate explanations [66]. As a result, it is fundamental to ensure that group members help each other mutually to perform the activities. Group objectives contribute to this inter-dependence.
- **Education stage and student profile.** Despite the fact that the evidence indicates that these strategies function well with all age groups and at every level of education, the effects are especially significant at primary school level [65] [18]. However, this is not to say that cooperative learning is necessarily ineffective at secondary level; the most plausible hypothesis is that its implementation in secondary school has seldom been the most appropriate. On the other hand, studies indicate that no significantly differential effects are shown depending on students initial competencies levels [20] [61]. Nevertheless, some cooperative learning programs appear to function especially well for students with greater learning difficulties [70].
- **Formal and structured learning activity.** Research agrees on the fact that an effective collaborative learning program requires a structured approach with well-designed activities [20]. That means that these programs have to incorporate curricular content, pedagogic methods and professional development [67].
- **Skills and disciplinary areas.** Activities, apart from having to be highly structured, must also place great emphasis on the basic instrumental competencies of the subject in question (maths, language, sciences, social sciences, etc.). In other words, studies show that cooperative groups have a greater impact on the science subjects, on subjects such as maths and sciences[65].
- **Incentives.** Evidence shows that collaboration can be encouraged by introducing some form of competitive approach between the groups [68] [56]. Obviously, this type of incentive is not always necessary and should be limited to avoid the focus of attention being placed excessively on the competition, instead of the learning. One example of a cooperative learning program which responds to this criteria is STAD (Student Teams-Achievement Divisions) [42] [62] [69].

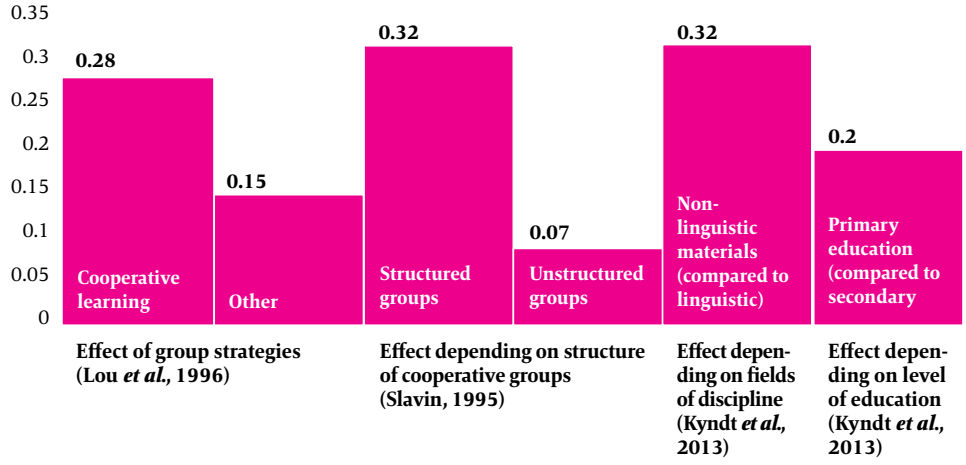
- **Combination with other methodological approaches.** Depending on the disciplinary subject in question, it might be necessary to integrate different methodological approaches into the cooperative learning method. For example, programs evaluated positively such as TAI [69] and BCIRC [49] integrate cooperative learning with individual work and direct instruction, respectively.

Graph 1.
Average effect of cooperative learning strategies according to discipline



Source: Prepared by authors based on meta-analysis referenced in graph.

Graph 2.
Differential effect of cooperative learning strategies



Source: Prepared by authors based on meta-analysis referenced in graph.

Box 2.

Two effective reading programs based on cooperative learning: the (Bilingual) Cooperative Integrated Reading and Composition program (CIRC & BCIRC), United States

CIRC is a cooperative learning program developed in 1983 by Robert Slavin and Nancy Madden at the Center for Social Organization of Schools of Johns Hopkins University. Later it became the programs *Reading Roots* (for early readers) and *Reading Wings* (for higher level primary school students). This program is one of the fundamental reading features of the comprehensive school reform model Success for All, which offers students the opportunity to practice their reading comprehension skills and reading in pairs or small groups.

The program is designed to assist students in developing meta-cognitive strategies to comprehend narrative and expositive texts. Once the teacher has explained the reading comprehension strategies, students work in groups of 4 or 5, in which the members establish a relationship of inter-dependency and mutual assistance. In these groups students read stories, predict how the story ends, prepare summaries and answer the teacher's questions. Within the cooperative teams, students work to comprehend the main ideas of the story and perform writing activities associated with the text. In these activities, students cover writing, spelling, decoding, vocabulary and reading fluency.

An adaptation of the CIRC is the Bilingual Cooperative Integrated Reading and Composition program (BCIRC), which is also based on the research evidence from second language acquisition. This program was designed to help Spanish-speaking students succeed in reading Spanish and then to make a successful transition to English reading.

The CIRC program has shown a potentially positive impact in reading comprehension among boys and girls in the early years of primary school, while it has shown no significant impact in reading performance [45] [47]. The impact in terms of reading comprehension seems to be maintained among adolescent students, who also obtained better scores in reading [48] [46]. On the other hand, implementation of the program's bilingual version has proved to be especially effective. Implementation of this program is shown to be positively associated with reading performance as well as in English language development [44] [49].

For further information:

Calderón, M., Hertz-Lazarowitz, R., & Slavin, R. E. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *Elementary School Journal, 99*(2), 153-165.

Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: effects on students' achievement, attitudes and social relations. *American Educational Research Journal, 32*(2), 321-351.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology, 83*(1), 8-16.

Table 4.
Cooperative learning programs & strategies. Meta-analyses reviewed

Meta-analyses (country of reference)	No stud. incl.	Research design	Duration of programs	Skills considered	Level of education	Average effects (AE) & differential effects (DE)	Effect measure-ment*
Lou *et al.* (1996) [18] (United States)	51	Exp. & qexp.*	3 terms: < 4 weeks 4-16 weeks > 17 weeks	Reading, maths & sciences	Primary, secondary or post-secondary	AE: Performance	+0.12
						DE: Cooperative learning	+0.28
						DE: Other group strategies	+0.15
Puzio & Colby (2013) [61] (United States)	18	Exp. & qexp.	From less than 10 weeks to 1	Language (reading, comprehension & vocabulary)	Primary & secondary lower-level	AE: Reading (n=16)	+0.16
						AE: Comprehension (n=18)	+0.20
						AE: Vocabulary (n=14)	+0.22
Nunnery *et al.* (2013) [62] (Uni. Kingdom)	15	Exp. & qexp.	Min. 12 weeks	Maths	Primary & secondary	AE: STAD program	+0.16
Kyndt *et al.* (2013) [65] (Belgium)	65	Qexp.	Short & long interventions & (n/a)	Performance (non-linguistic proficiency: sciences, maths; linguistic proficiency: language, social sciences), attitudes & perceptions	Primary, secondary & tertiary	AE: Performance	+0.54
						DE: Performance (non-linguistic *vs.* linguistic proficiency)	+0.32
						DE: Performance (primary ed. vs. secondary ed.)	+0.20
						DE: Performance (tertiary ed. vs. secondary ed.)	+0.18
						AE: Attitudes	+0.15
						AE: Perceptions	+0.18 (ns)
Romero (2009) [63] (United States)	30	Exp. & qexp.	(n/a)	Sciences	Secondary & post-secondary	AE: Performance	+0.31
Stoner (2004) [64] (United States)	25	Exp. & qexp.	From 4 weeks to 1 academic year	Maths	Secondary & lower grades	AE: Performance	+0.14
Slavin (1995) [20] (United States)	99	Exp. & qexp.	Min. 4 weeks	Reading & maths	Secondary	DE: Structured groups*** (n=64)	+0.32
						DE: Non-structured groups (n=35)	+0.07(ns)
Slavin, Lake, Chambers *et al.* (2009) [67] (United States)	10	Exp. & qexp.	From 6 months to 2 years	Reading	Primary	AE: Program CIRC (n=9)	+0.21
Slavin & Lake (2008) [69] (United States)	9	Exp. & qexp.	From 12 weeks to 2 years	Maths	Primary	AE: STAD program (n=4)	+0.30
						AE: TAI program: cooperative individualized learning (n=5)	+0.20
Slavin *et al.* (2008) [40] (United States)	11	Exp. & qexp.	From 12 weeks to 3 years	Reading	Secondary	AE: Cooperative learning strategies (n=7)	+0.28
						EM: Student Team Reading Reading Edge programs (n=4)	+0.29
Slavin, Lake & Groff (2009) [56] (United States)	7	Exp. & qexp.	1 year (except 1 of 18 weeks & another 1 semester)	Maths	Secondary	AE: Cooperative learning strategies	+0.46
						DE: STAD program (n=4)	+0.42
						DE: IMPROVE program(n=3)	+0.52

Source: Prepared by authors based on Lou *et al.* (1996), Puzio & Colby (2013), Nunnery *et al.* (2013), Kyndt *et al.* (2013), Romero (2009), Stoner (2004), Slavin (1995), Slavin, Lake, Chambers *et al.* (2009), Slavin & Lake (2008), Slavin *et al.* (2008) & Slavin, Lake & Grof (2009).
* Qexp.: quasi-experimental. Exp.: experimental.
** Standardized effect size (Cohen,1988). Small effect: 0.2. Medium effect: 0.5. Large effect: 0.8.
*** Structured Team Learning Methods are cooperative learning methods that offer some form of award based on the sum of members' individual learning.
n/a: Not available.

Summary

The evidence reviewed, in the first place, shows us that both homogenous as well as heterogeneous grouping has an average positive effect on learning, more so than teaching the entire class group. This positive impact is demonstrated both as far as cognitive outcomes are concerned as well as non-cognitive skills are concerned. The effects of heterogeneous grouping would be equivalent to a half school year of academic progress. All students, especially those with lower learning abilities, benefit from this type of grouping.

Secondly, the evidence indicates that homogeneous grouping between classes has practically no effect on academic performance. There is a lack of consensus regarding the differential effects of this type of grouping. Even though, in general, the evidence cannot identify whether they hold more benefits for a certain student profile, there is a large quantity of research available to indicate that students in higher level groups benefit from tracking between classes, while the opposite is true for students in groups of lower academic values.

By contrast, homogeneous grouping within classes does seem to favor medium learning levels. This effectiveness can also be observed in homogeneous multi-grade and multi-age grouping, above all when performed for one or two subjects. The positive impact appears to be associated with a high level of formal organizational flexibility of the groups, a fact which enables students' mobility between groups and is combined with instruction in a heterogeneous class group. The effects of homogeneous grouping are positive for all students when compared with class group instruction (in which disadvantaged students are more vulnerable). However, if the comparison is made with heterogeneous grouping strategies, ability tracking or streaming appear to have a detrimental effect on lower skills level students and to favor students with average and higher academic levels.

Finally, heterogeneous grouping strategies based on cooperative learning have a positive impact on students' performance insofar as they are formal and well-structured: there are group objectives, individual accountability and reward or acknowledgement system for task completion. The impact is evidenced in cognitive as well as non-cognitive outcomes (attitudes and perceptions). These strategies function well for any students, independently of their academic level, at all education levels (especially in primary school) and for all subjects (especially scientific subjects).

Table 5.
Arguments for and against grouping modalities

For	Against
Student grouping strategies	
Positive impact on learning, as differentiated instruction strategies are carried out, additional materials are used and teaching staff are trained in their administration.	Grouping for the sake of grouping does not produce any significant effects. Certain minimum conditions must be guaranteed to ensure additional resources (materials, training, etc.) which are not always available.
Positive impacts verify the cognitive as we as non-cognitive outcomes.	
Ability tracking or streaming between classes	
The definition of more specific learning objectives allows for increasing teaching time and offer a greater quantity of instruction.	No average significant effects of this kind of grouping in learning have been identified.
The cost of implementing grouping by levels between classes is low.	The effects for more academically vulnerable students are not clear. Some research indicates negative effects.
	Overly rigid structure does not allow for student mobility between groups depending on their academic progress: perpetuating initial inequality.
Ability tracking or streaming within classes	
In comparison with heterogeneous grouping and traditional teaching, they have an average positive impact on learning.	In comparison with heterogeneous grouping, they can be detrimental to students in homogeneous groups with a lower academic value.
In comparison with traditional teaching, they have a positive impact for all students, including those with a low level of skills.	Ability tracking or streaming are not effective if teaching methods and materials are not adapted.
Defining more concrete learning objectives permits expanding teaching time and offering a greater quantity of instruction.	Evaluation systems are required for ongoing monitoring of students' progress thereby guaranteeing mobility between groups. This requires time and resources.
Grouping within class can be quite flexible to ensure students' mobility between groups in accordance with their academic progress.	Grouping for the sake of grouping does not produce any significant effect. Certain minimum conditions must be guaranteed to ensure additional resources (materials, training, etc.) which are not always available.
Heterogeneous grouping based on cooperative learning	
In comparison with homogeneous grouping and traditional teaching, they have an average positive impact on learning.	Cooperative grouping are not effective when structured in an improvised manner, are informal or unstructured.
Benefit all students if they are formals, well-structured and applied regularly.	Grouping for the sake of grouping does not produce any significant effect. Certain minimum conditions must be guaranteed to ensure additional resources (materials, training, etc.) which are not always available.
Facilitate social cohesion in class: generate positive inter-dependence and a relationship of mutual assistance between students.	
The cost of implementing cooperative learning programs is low.	

Implications for practice

According to the empirical evidence reviewed, we can establish a hierarchy of "convenience" in the application of grouping: flexible ability grouping strategies by levels organized within the class group are preferable to ability tracking or streaming between classes. In this way, the preferred option is heterogeneous grouping, organized by way of cooperative learning strategies, with regard to flexible grouping by levels within the class group. Even though flexible grouping by levels may be a more effective means than heterogeneous grouping, the detriment produced for students of a lower academic level can be much higher.

Centers where flexible grouping strategies are implemented must allow for student mobility from one group to another.

So, what are the implications of this conclusion in practice? During the last thirty years, in Catalonia there have been several debates about how best to deal with diversity by introducing individualized and personalized teaching proposals. For example, the introduction of the General Organization of the Education System Act of 1990 (*Ley Orgánica de Ordenación General del Sistema Educativo* or *LOGSE*); generated a debate about the benefits of adapting the curriculum to suit students with specific educational needs or those with either permanent or temporary learning difficulties. Later on, another debate arose concerning the introduction of *aules d'acollida* (newcomers' program) as a resource to facilitate the incorporation of recently arrived students joining regular class groups. Currently, there are other measures being applied to address diversity such as the Personalized Education Support and the Special Education Support Units. However, the most controversial measure to date, especially among families, has been the flexible grouping strategies, which were implemented in the majority of centers by grouping students by levels in a homogeneous fashion.

The empirical evidence supports this practice. That said, it supports it whenever certain basic working conditions are adhered to:

- Centers where flexible grouping strategies are implemented must allow for student mobility from one group to another. Otherwise, grouping can become an inefficient measure which strengthens existing educational inequalities that affect the more vulnerable students. This mobility can only be achieved by establishing a systematic and ongoing evaluation system of the students' progress.
- The second condition is that teaching methods and materials used must be adapted to the different levels of the varying groups: Not only is this a required condition for addressing the different needs of students, but also to accelerate learning pace: aligning the more specific learning methods and defining objectives allows permits teachers to increase teaching time and offer a greater quantity of instruction.
- The third condition is that these strategies do not translate into a measure resembling grouping by levels between classes. This can be avoided when grouping is only used in one or two fields of discipline, such as for example, linguistics and maths, as is the case today in many centers

The evidence corroborates the importance of cooperative grouping, but also prescribes certain conditions to ensure that cooperative strategies have a positive impact on learning.

On the other hand, Catalonia has a long-standing pedagogical tradition that has enabled many teachers to utilize heterogeneous grouping to deal with diversity. This samediversity is functional for organizing teaching and learning processes. In the context of heterogeneous grouping, for many years now, educators have been experimenting with cooperative learning methods of greater and lesser formality. In fact, many of the efforts focusing on managing diversity by way of cooperative groups have translated into tangible improvements for many academic centers, above all by virtue of a significant investment in time and motivation. Even so, other initiatives turned out to be no more than well intended efforts but not systematic enough to overcome a traditional teaching model incapable of coping with students' heterogeneity.

The research evidence corroborates the importance of this method, but also prescribes certain conditions to ensure that cooperative strategies have a positive impact on learning. The principal condition is to ensure well-designed and well-structured grouping practices:

- When we refer to well-structured groups we are referring, fundamentally, to the design of group and individual incentives: as indicated above, cooperative groups must have a well-defined group objective which structures the activity.
- Every member in the group must be responsible for their own task, contributing individually to the group objective.
- Moreover, groups tend to function better if there are collective and individual rewards associated with meeting group objectives.

Ultimately, significant changes are produced when systematic, sequential and planned action is taken which incorporates cooperative grouping methods into the center's regular methodology.

Bibliography

[1] Smith, M. L. & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal, 17* (4), 419-433.

[2] Krueger, A. B. & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: evidence from Project STAR. *The Economic Journal, 111* (468), 1-28.

[3] Nye, B. & Hedges, L. V. (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.

[4] 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.

[5] Nye, B., Hedges, L. V. & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: the results of the Tennessee class size experiment. *American Educational Research Journal, 37* (1), 123-151.

[6] Whitehurst, G. J. & Chingos, M. M. (2011). Class size: what research says and what it means for state policy. May 11. Brown Center on Education Policy at the Brookings Institution.

[7] Angrist, J. D. & Lavy, V. (1999). Using Maimonides' Rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics, 114* (2), 533-575.

[8] Bressoux, P., Kramarz, F. & Prost, C. (2008). Teachers' training, class size and students' outcomes: learning from administrative forecasting mistakes. IZA Discussion Paper, 3871. Institute for the Study of Labor (IZA).

[9] Fredriksson, P., Öckert, B. & Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics, 128* (1), 249–285.

[10] Duflo, E., Dupas, P. & Kremer, M. (2012). School governance, teacher incentives, and pupil-teacher ratio: experimental evidence from Kenyan primary schools. NBER Working Paper, 17939. National Bureau of Economic Research.

[11] Chingos, M. M. (2012). The impact of a universal class-size reduction policy: evidence from Florida's statewide mandate. Economics of Education Review, 31 (5), 543-562.

[12] Jepsen, C. (2015). Class size: does it matter for student achievement? IZA World of Labor. Institute for the Study of Labor (IZA). Recovered from http://wol.iza.org/articles/class-size-does-it-matter-for-student-achievement

[13] Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73* (2), 417-458.

[14] Hoxby, C. M. (2000). The effects of class size on student achievement: new evidence from population variation. *The Quarterly Journal of Economics, 115* (4), 1239-1285.

[15] Betts, J. R. & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: the case of math teachers. *Educational Evaluation and Policy Analysis, 21* (2), 193-213.

[16] The Education Endowment Foundation. (2015). *Reducing class size.* https://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/reducing-class-size

[17] Puzio, K. & Colby, G. T. (2010). *The effects of within class grouping on reading achievement: a meta-analytic synthesis.* Evanston, IL: Society for Research on Educational Effectiveness.

[18] Lou, Y., Abrami, P. C., C., S. J., Poulsen, C., Chambers, B. & d'Apollonia, S. (1996). Within-class grouping: a meta-analysis. *Review of Educational Research, 66* (4), 423-458.

[19] 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.

[20] Slavin, R. E. (1995). *Co-operative learning: theory, research, and practice* (Ed. 2). Boston: Allyn and Bacon.

[21] Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: a best-evidence synthesis. *Review of Educational Research, 57* (3), 293-336.

[22] Slavin, R. E., Daniels, C. & Madden, N. A. (2005). "Success for All" middle schools add content to middle grades reform. *Middle School Journal, 36* (5), 4-8.

[23] Slavin, R. E., Chamberlain, A. M., Daniels, C. & Madden, N. A. (2009). The Reading Edge: a randomized evaluation of a middle school cooperative reading program. *Effective Education, 1* (1), 13-26.

[24] Chamberlain, A. M., Daniels, C., Madden, N. A. & Slavin, R. E. (2007). A randomized evaluation of the Success for All middle school reading program. *Middle Grades Research Journal, 2* (1), 1-21.

[25] What Works Clearinghouse. (2012b). *Reading Edge.* Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_readingedge_062612.pdf

[26] Tobin, K. G. (2004). The effects of beginning reading instruction in the Horizons Reading program on the reading skills of third and fourth graders. *Journal of Direct Instruction, 4* (2), 129-137.

[27] CSRQ. (2006). *CSRQ Center report on elementary school comprehensive school reform models*. Washington, DC: Comprehensive School Reform Quality Center. American Institutes for Research.

[28] Biancarosa, G., Bryk, A. S. & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal, 111* (1), 7-34.

[29] Kulik, J. A. & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36* (2), 73-77.

[30] Kulik, J. A. (1992). *An analysis of the research on ability grouping: historical and contemporary perspectives*. Storrs, CT: The National Research Center on the Gifted and Talented (University of Connecticut).

[31] Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: a best-evidence synthesis. *Review of Educational Research, 60* (3), 471-499.

[32] Gutiérrez, R. & Slavin, R. E. (1992). Achievement effects of the nongraded elementary school: a best evidence synthesis. *Review of Educational Research, 62* (4), 333-376.

[33] Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51* (6), 842-858.

[34] Jackson, C. K. (2009). Ability-grouping and academic inequality: evidence from rule-based student assignments. NBER Working Paper, 14911. National Bureau of Economic Research.

[35] Argys, L. M., Rees, D. I. & Brewer, D. J. (1996). Detracking America's schools: equity at zero cost? *Journal of Policy Analysis and Management, 15* (4), 623-645.

[36] Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14* (3), 205-227.

[37] Betts, J. R. & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review, 19* (1), 1-15.

[38] Robinson, J. P. (2008). Evidence of a differential effect of ability grouping on the reading achievement growth of language-minority hispanics. *Educational Evaluation and Policy Analysis, 30* (2), 141-180.

[39] Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review, 57* (6), 812-828.

[40] Slavin, R. E., Cheung, A. C. K., Groff, C. & Lake, C. (2008). Effective reading programs for middle and high schools: a best-evidence synthesis. *Reading Research Quarterly, 43* (3), 290-322.

[41] Johnson, D. W., Johnson, R. T. & Holubec, E. J. (1999). *El aprendizaje cooperativo en el aula*. Buenos Aires: Paidós.

[42] Slavin, R. E., Sharan, S., Kagan, S., Hertz-Lazarowitz, R., Webb, C. & Schmuck, R. (Eds.). (1985). *Learning to cooperate, cooperating to learn*. Nova York: Springer Science+Business Media.

[43] Johnson, D. W. & Johnson, R. T. (1992). Positive interdependence: key to effective cooperation. A R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: the theoretical anatomy of group learning*. New York: Cambridge University Press.

[44] Calderón, M., Hertz-Lazarowitz, R. & Slavin, R. E. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *Elementary School Journal, 99* (2), 153-165.

[45] Stevens, R. J., Slavin, R. E. & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology, 83* (1), 8-16.

[46] Stevens, R. J. & Slavin, R. E. (1995). The cooperative elementary school: effects on students' achievement, attitudes and social relations. *American Educational Research Journal*, 32 (2), 321-351.

[47] What Works Clearinghouse. (2012a). *Cooperative Integrated Reading and Composition (beginning readers)*. Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_circ_062612.pdf

[48] What Works Clearinghouse. (2010). *Cooperative Integrated Reading and Composition (adolescent learners)*. Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_circ_083110.pdf

[49] What Works Clearinghouse. (2007). *Bilingual Cooperative Integrated Reading and Composition*. Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_BCIRC_021507.pdf

[50] Slavin, R. E., Leavey, M. B. & Madden, N. A. (1984). Combining cooperative learning and individualized instruction: effects on student mathematics achievement, attitudes, and behaviors. *The Elementary School Journal, 84* (4), 409-422.

[51] Stevens, R. J. & Durkin, S. (1992). Using student team reading and student team writing in middle schools: *Two evaluations. Part II. Student team reading and student team writing: An evaluation of a middle school reading and writing program*. Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students.

[52] Stevens, R. J. (2003). Student team reading and writing: a cooperative learning approach to middle school literacy instruction. *Educational Research and Evaluation, 9* (2), 137-160.

[53] What Works Clearinghouse. (2011). *Student Team Reading and Writing*: Institute of Education Sciences, WWC Intervention Report. http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_str_111511.pdf

[54] Suyanto, W. (1998). *The effects of student teams-achievement divisions on mathematics achievement in Yogyakarta rural primary schools*. University of Houston, Houston, TX.

[55] Mevarech, Z. R. & Kramarski, B. (1997). Improve: a multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34* (2).

[56] Slavin, R. E., Lake, C. & Groff, C. (2009). Effective programs in middle and high school mathematics: a best-evidence synthesis. *Review of Educational Research, 79* (2), 839-911.

[57] Schoen, H. L., & Hirsch, C. R. (2002). The Core-Plus Mathematics project: perspectives and student achievement. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: what are they? what do students learn?* (pp. 311-343). Hillsdale, NJ: Lawrence Erlbaum Associates.

[58] Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N. & Phillips, E. D. (1998). *Connected Mathematics*. White Plains, NY: Dale Seymour Publications.

[59] Fendel, C., Resek, D., Alper, L. & Fraser, S. (1996). *Interactive mathematics program (IMP)*. Berkley, CA: Key Curriculum Press.

[60] Lago, J. R. & Pujolàs, P. (2012). Un programa para cooperar y aprender. *Cuadernos de pedagogía, 428*, 24-26.

[61] Puzio, K. & Colby, G. T. (2013). Cooperative learning and literacy: a meta-analytic review. Journal of Research on Educational Effectiveness, 6 (4), 339-360.

[62] Nunnery, J. A., Chappell, S. & Arnold, P. (2013). A meta-analysis of a cooperative learning model's effects on student achievement in mathematics. *Cypriot Journal of Educational Sciences, 8* (1), 34-48.

[63] Romero, C. C. (2009). *Cooperative learning instruction and science achievement for secondary and early post-secondary students: a systematic review*. Colorado State University, Fort Collins, CO.

[64] Stoner, D. A. (2004). *The effects of cooperative learning strategies on mathematics achievement among middle-grades students: a meta-analysis*. University of Georgia, Athens.

[65] Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E. & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review, 10*, 133-149.

[66] Webb, N. M. (1985). Student interaction and learning in small groups. A R. E. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb & R. Schmuck (Eds.), *Learning to cooperate, cooperating to learn*. Nova York: Springer Science+Business Media.

[67] Slavin, R. E., Lake, C., Chambers, B., Cheung, A. C. K. & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79* (4), 1391-1466.

[68] Reid, J. (1992). The effects of cooperative learning with intergroup competition on the math achievement of seventh grade students. ERIC Document Reproduction Service No. ED 355106.

[69] Slavin, R. E. & Lake, C. (2008). Effective programs in elementary mathematics: a best-evidence synthesis. *Review of Educational Research, 78* (3), 427-515.

[70] Bramlett, R. K. (1994). Implementing cooperative learning: a field study evaluating issues for school-based consultants. Journal of School Psychology, 32(1), 67-84.

First edition: November, 2015
© Fundació Jaume Bofill, Ivàlua, 2015
fbofill@fbofill.cat, info@ivalua.cat
http://www.ivalua.cat
http://www.fbofill.cat

Authors: Miquel Àngel Alegre, Gerard Ferrer-Esteban
Edited by: Fundació Jaume Bofill
Design & layout: Enric Jardí
ISBN: 978-84-944534-6-5
DL: B 3949-2016

With the support of:
Obra Social "la Caixa"

This work is subject to the Creative Commons license **Attribution-Non-commercial-NoDerivs (BY-NC-ND)**. You can reproduce, distribute and publicly communicate this work whenever you attribute authorship. You may not make commercial use or produce derivatives.