Educational Research Review 21 (2017) 33–54

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Review

Academic achievement of students without special educational needs in inclusive classrooms: A meta-analysis

Grzegorz Szumski*, Joanna Smogorzewska, Maciej Karwowski

The Maria Grzegorzewska University, Department of Educational Sciences, Szczesliwicka St., 40, 02353 Warsaw, Poland

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 14 February 2017
Accepted 23 February 2017
Available online 7 March 2017

Keywords:
Inclusive education
Academic achievement
Students without special educational needs
Meta-analysis

ABSTRACT

This article presents a meta-analysis that attempts to establish how the presence of students with special educational needs in the classroom impacts students without special educational needs. We meta-analyzed 47 studies that met the inclusion criteria, covering a total sample of almost 4 800 000 students. The overall effect was positive and statistically significant but weak, $d = 0.12$ (95% CI: 0.02, 0.23). A number of moderators, including the country of study, the manner of implementation (intervention studies versus regular school practice), the educational team composition, the level and type of disorders in students with special educational needs, and finally educational stage were examined. We discuss the findings in terms of assumptions and controversies surrounding the very concept of inclusive education.

Contents

1. Introduction 34
 1.1. Inclusive education: positions and controversies 34
 1.2. Previous reviews on the academic achievement of students without SEN in inclusive classrooms 35
 1.3. Factors that may moderate the academic achievement of students without SEN in inclusive classrooms 36
  1.3.1. Country of the study 36
  1.3.2. Inclusive education implementation 36
  1.3.3. Educational team composition 36
  1.3.4. Students with severe versus mild SEN in the classroom 37
  1.3.5. Inclusive classrooms with and without students with EBD 37
  1.3.6. Educational stage 37
  1.3.7. Proportion of students with SEN in the classroom 38
  1.3.8. Preparation of general education teachers 38
  1.3.9. Co-teaching models 38
2. Method 38
 2.1. Search strategies 38
 2.2. Criteria for including and excluding studies 39
 2.3. Assessment of the quality of studies 40
 2.4. Data coding 40
 2.5. Moderators 40

* Corresponding author.
E-mail address: bgszum@poczta.onet.pl (G. Szumski).

http://dx.doi.org/10.1016/j.edurev.2017.02.004
1747-938X/

2.6. Statistical analyses . . . 43
3. Results . . . 43
3.1. Publication bias . . . 43
4. Discussion . . . 47
4.1. Academic achievement of students without SEN and the legitimization of inclusive education . . . 47
4.2. Academic achievement of students without SEN and inclusive education concept . . . 47
4.3. Support from special education teachers and the academic achievement of students without SEN . . . 48
4.4. Students' type of SEN and the academic achievement of their classmates without SEN . . . 48
4.5. Educational stage and students' academic achievement . . . 49
5. Limitations and future studies . . . 49
6. Conclusion . . . 49
Acknowledgement . . . 50
References . . . 50

1. Introduction

This article explores one of the crucial problems in contemporary discourses on inclusive education: to what extent do students without special educational needs (SEN) who are taught together with classmates with SEN, achieve academically more or less than when taught in homogeneous classes. This question is among the main issues raised by both proponents (Ainscow, Dyson, Goldrick, & West, 2012; Lipsky & Gartner, 1996) and skeptics (Hornby, 2011; Kauffman & Hallahan, 2005) of inclusive education. Our investigation has two primary aims. The first is to provide an average meta-analytically obtained effect size, describing the direction and strength of the effect of the presence of students with SEN on the school achievement of their peers without SEN. The second aim is to examine the potential impact of theoretically derived factors that may moderate this effect. Although there are examples of previous systematic reviews that have addressed related issues (e.g., Kalambouka, Farrell, Dyson, & Kaplan, 2007; Ruijs & Peetsma, 2009), we are not aware of any published meta-analytic synthesis of this problem. We therefore endeavor to shed light on the mixed findings of prior research by providing a more stable estimate of the impact on school achievement in the presence of students with SEN on their peers without SEN and by examining factors that may potentially moderate this relationship. We begin this article by briefly defining inclusive education and discussing the main controversies raised in discourses on inclusive education. We then review the main theoretical and empirical positions regarding the school achievement of students with and without SEN. Finally, we discuss the role of potential moderators that may be responsible for the different results obtained in many of the studies.

1.1. Inclusive education: positions and controversies

Inclusive education is an ambiguous term. Broadly, it refers to a situation in which students with SEN are taught in mainstream schools and spend at least some of the time in the classroom together with their peers without SEN (Farrell, Dyson, Polat, Hutcheson, & Gallannaugh, 2007). Narrower interpretations add criteria that relate to educational organization processes, support for students with diverse needs, and sometimes also the quality of SEN students' participation in school as well as to educational arrangements that are optimal for all students (Booth & Ainscow, 2002). These narrower interpretations inform our inquiry here, as they stress the potential of inclusive education to improve school as a place where all students learn.

Inclusive education is usually recommended, not only as a direction for changing the education of people with disabilities (see: UN, 2006, Article 24), but also as a new model for school, specifically responding to the diverse needs of contemporary societies and being consistent with the democratic values on which these societies are built (Artiles, Kozleski, Dorn, & Christstensen, 2006; Opertti, Walker, & Zhang, 2014). An approach focusing on global school reform assigns inclusive education many more functions than the traditional approach, recently called 'inclusive special education' (Hornby, 2014). Its adherents claim that inclusive education is a never-ending movement of restructuring climate, policy and school practices, to prevent any student from being excluded from culture, programs and society (Ainscow, Booth, & Dyson, 2006). Inclusion understood as a "principled approach to education and society" (Ainscow et al., 2006, p. 22) is not limited to a narrow understanding of school aims, typical for the "market" (Miles & Singal, 2010), such as academic achievement. Indeed, individual and social development of students includes not only their academic achievement, but also a sense of well-being and proper social relations with peers and teachers (Loreman, 2014; Soukakou, 2012).

The practical impact of inclusive education on educational policy and practice stems above all from the ethical and legal arguments used by its advocates. On the other hand, many scholars stress the shortage and limited quality of empirical studies on the effectiveness of inclusive education and its determinants (Farrell, 2000; Göransson & Nilholm, 2014; Lindsay, 2007). These remarks are particularly valid when it comes to analyses of the school achievement of students without SEN. Even though such students constitute a majority in inclusive classrooms, studies devoted to inclusive education usually focus

on students with SEN (Freeman & Alkin, 2000; Lindsay, 2007). Integrating the perspectives on rights and ethics (i.e. Lipsky & Gartner, 1996; Slee, 2001) with the perspective of efficacy (i.e. Lindsay, 2007), may enrich scholarly discourse on inclusive education and provide relevant arguments in the debate for its promotion (Artiles, Harris-Murri, & Rostenberg, 2006). What makes such conceptual integration possible is liberal theories of social justice, especially theorizing justice as fairness (Rawls, 1971). According to Rawls, two principles should be taken into account when considering the justice of a social institution. First, public institutions should treat all citizens equally; second, unequal treatment is justified only if the weakest members of society benefit. Excluding students with SEN from access to general education may be regarded as unjust (Miles & Singal, 2010) because such exclusion restricts their access to the common good. On the other hand, however, this argument cannot justify an inclusive education system as such. It fails to consider whether an inclusive school ensures that students with SEN attain at least as good an outcome as they would in a special school (Freeman & Alkin, 2000; Oh-Young & Filler, 2015). Moreover — which is important to our analyses — it fails to justify whether or not inclusive education harms the outcomes of students without SEN (Connor & Ferri, 2007). Rawls' second principle may be violated if students without SEN in inclusive classrooms have poorer achievement than in regular classrooms.

Although some of the previous analyses suggest that inclusive education has a beneficial influence on the academic achievement of students without SEN, this effect may be moderated by the manner of implementation of inclusive education (Ainscow et al., 2012; Giangreco & Suter, 2015). The admission of children with SEN rarely leads to changes in the organization of school, its curriculum, and teaching and learning strategies (Ainscow & Cesar, 2006). Consequently, it does not always lead to an improvement in students' academic achievement.

However, inclusive education can also be understood as a transformative approach to education, which reforms educational processes and organization (Ainscow & Cesar, 2006; Giangreco & Suter, 2015). These changes may be introduced both at the school level and at the classroom level, for example by creating more effective teaching strategies (Erten & Savage, 2012). Advocates of the former approach assume that the potential of inclusive education lies primarily in the change it brings to the philosophy and organization of school, comprising inclusive policies, cultures, and practices (Booth & Ainscow, 2002). Advocates of the other orientation assume that the inclusion of students with SEN promotes more effective teaching strategies in the school (e.g., Bottge et al., 2015). Both mechanisms may likely have a positive impact on students' outcomes, and there may be a symbiotic relationship between school effectiveness and effective teaching (Taddlie, Kirby, & Stringfield, 1989). Strategies of adapting instruction for students with SEN, such as: control of task difficulty, teaching in small collaborative groups, and directed response questioning, also work effectively for learners without SEN (Vaughn, Gersten, & Chard, 2000). Similarly, instructional strategies such as frequent feedback, cooperative learning, focus on concepts, or positive classroom climates and sensitive teachers, whose effectiveness has been confirmed in numerous studies and meta-analyses (Hattie, 2009), also work effectively in inclusive classrooms (Mitchell, 2014).

Importantly, changes on a school level can improve students' school achievement as well. Rouse and Florian (2006) have shown that average school achievement in schools with higher percentage of students with SEN is comparable to average achievement in schools without such students. This study has also suggested that implementing inclusive practices on school level can lead to achievement improvement among all students when the growth of academic achievement among students, and not the fulfillment of a country's standards, is compared. Similarly, Ainscow and Cesar (2006) have demonstrated that development of inclusive practices in schools is linked to the improvement of students' results in achievement tests. However, even this result tells us little about possible causality.

In educational reality schools tend to reactively adapt to the increasing number of students with SEN rather than proactively plan and re-organize a school in such a way as to make it ready for the effective education of heterogeneous groups (Giangreco & Suter, 2015). Obviously, creating an effective school for all students is difficult since it requires cooperation with other schools and the broader community, team decision making, flexible use of resources, and a new system for the preparation of teachers and special educators (Ainscow et al., 2012; Causton & Theoharis, 2014; Cusumano, Algozzine, & Algozzine, 2014).

1.2. Previous reviews on the academic achievement of students without SEN in inclusive classrooms

Two recent systematic reviews of the effects of inclusive education for school achievement of students without SEN (Kalambouka et al., 2007; Ruijs & Peetsma, 2009) were published almost simultaneously. The paper by Kalambouka et al. (2007) is devoted exclusively to the outcomes of students without SEN. Kalambouka reviewed 26 studies of which 21 were devoted to academic outcomes. Positive effects were found in six studies, negative effects in one study, and the remaining studies demonstrated that attending inclusive classrooms is unrelated to the academic achievement of students without SEN. Kalambouka and colleagues also tried to estimate the impact of the type of SEN and educational stage on the academic achievement of students without SEN, yet they did not find any relationship between the type of SEN and classmates' academic achievement.

The review by Ruijs and Peetsma (2009) is mostly based on the same studies as the analysis by Kalambouka et al. (2007), though the authors included several studies published after 2003 — a period not covered by older reviews (Cole, Waldron, & Majd, 2004; Demeris, Childs, & Jordan, 2007; Dyson, Farrell, Polat, Hutcheson, & Gallannaugh, 2004; Rouse & Florian, 2006). Three of these studies found a positive effect of attending inclusive classrooms (Cole et al., 2004; Demeris et al., 2007; Rouse & Florian, 2006), while in one study this effect was negative (Dyson et al., 2004). Ruijs and Peetsma (2009) concluded that their

review led to ambiguous conclusions about the impact of inclusive education on the academic achievement of students without SEN.

These ambiguous conclusions of previous reviews informed our decision to conduct a quantitative synthesis: meta-analysis to estimate the effect size of the relationship between inclusive education and academic achievement of students without SEN.

1.3. Factors that may moderate the academic achievement of students without SEN in inclusive classrooms

On top of different ways of understanding inclusive education and many ways to implement it, several factors may qualify the size of differences in the achievement of students without SEN in inclusive versus regular classrooms. In the following section, we focus on the 9 factors that serve as theoretically relevant moderators. We included six of them (country of the study, inclusive education implementation, educational team composition, students' level and type of SEN, and educational stage) into our meta-analysis, presented below. Including the three other moderators (proportion of students with SEN in the classroom, preparation of general education teachers [GETs], and co-teaching model) was impossible due to the lack of appropriate data.

1.3.1. Country of the study

The country may play a role in explaining the effectiveness of inclusive education for several reasons: the length of experience in the implementation of inclusive education, the consistency of educational policy promoting inclusion, and the way inclusion is understood. All these factors make it legitimate to hypothesize that inclusion may bring more positive results in the U.S. and Canada than in continental Europe. Whereas in the U.S. the Education for All Handicapped Children Act (1975) guaranteed access to local public schools to all children with disability as a civil right, most European countries did not adopt similar laws until the 1990s (Powell, 2009). Furthermore, in many European countries laws permit, but do not guarantee, the admission of students with SEN to inclusive schools (see Emanuelsson, Haug, & Persson, 2005; Powell, 2009). Furthermore, due to early tracking in most European countries, particularly strong in Germany, Switzerland, and Austria (European Commission, 2014), teachers have less experience in teaching heterogeneous groups of learners than do teachers in North America (Kozleski, Artiles, & Waitoller, 2014). Finally, whereas in the U.S. and Canada inclusion is understood as a process of restructuring general education, in most European countries access to schools is granted to some students with SEN (Norwich, 2008). Therefore, the broadening of access to general schools in continental Europe is, above all, associated with efforts to ensure special educational support for students with SEN and less so with efforts towards a restructuring of teaching in the classroom and the school. The English educational system is an exception, but it is also ambiguous. According to the law, the aim of inclusive education in England is not only to address special educational needs of children with statements, but also to remove barriers in learning for all students. However, the educational policy in England also creates strong mechanisms which block disseminating inclusive education understood as restructuring schools to be more responsive to the diversity of students (Ainscow et al., 2006; Armstrong & Barton, 2008).

1.3.2. Inclusive education implementation

Research on the effects of inclusive education usually comprises two types of studies: (a) evaluations of interventions — educational experiments, implementations, or pilot studies, and (b) studies of regular school practice. Two arguments lead us to hypothesize that the academic achievement of students without SEN who were educated in experimental schools or schools implementing intervention projects is higher than those of regular school practice. First, in interventions, participants receive special attention from their observers and supervisors (Adair, Sharpe, & Huynh, 1989). Second, if inclusive education is equated with a change in the school or even a change of the whole school system, it should be implemented at least at the school district level (Hoppey & McLeskey, 2013). Because this kind of change is difficult, teachers have to receive support on the optimal use of resources and on improving their competencies (Erten & Savage, 2012; Loreman, 2007; Stanovich & Jordan, 1998). Such support that increases their sense of security and gives them an opportunity to improve their skills (McLeskey & Waldron, 2002; McLeskey, Waldron, & Reed, 2014) is usually provided precisely during the implementation of programs to transform schools into inclusive institutions.

1.3.3. Educational team composition

Debates on educational team composition in inclusive education concern team size (Zigmond, 2006) and qualifications of team members working in the classroom (Giangreco & Broer, 2005). In some countries, for example, in the Netherlands, the United Kingdom, and Switzerland, the practice is still to include individual students with higher-incidence disabilities (e.g., mild intellectual disabilities [MID] or emotional and behavioral disorders [EBD]) in general classrooms without additional support, or with support from a special education teacher (SET) for only a limited number of hours each week (Sermier Dessemontet & Bless, 2013; van der Veen, Smeets, & Derriks, 2010). Moreover, in some countries, especially in the USA, additional support in the classroom often entails paraprofessionals without proper qualifications in special education (Giangreco, Broer, & Suter, 2011). Both these solutions, i.e., no support or support of paraprofessionals, were found to be ineffective regarding students with SEN (Ainscow et al., 2012; Giangreco & Broer, 2005). We hypothesize it may also be less effective than full-time co-teaching between GET and SET in the case of students without SEN. General education teachers are poorly prepared to work in inclusive education classrooms (Forlin, 2001) and their low self-efficacy (Peebles & Mendaglio,

2014) may limit the effectiveness of instruction. The lack of support from a special education co-teacher was identified as a factor that intensifies stress and occupational burnout (Talmor, Reiter, & Feigin, 2005), which can also contribute to a decrease in the effectiveness of instruction. Although paraprofessionals may help GETs by giving them more time to work with students without SEN, they usually do not possess sufficient knowledge about effective learning methods to be used in heterogeneous classrooms (Giangreco, Suter, & Doyle, 2010).

Although the number of studies on the effectiveness of co-teaching in inclusive classrooms is insufficient, it can be hypothesized that co-teaching between GETs and SETs does have some positive effect on students' academic achievement (Solis, Vaughn, Swanson, & McCulley, 2012). Future studies, however, are necessary to explore this line of research.

1.3.4. Students with severe versus mild SEN in the classroom

There are at least three reasons to hypothesize that academic achievement of students without SEN taught in inclusive classrooms with students with severe SEN will be lower than achievement of their peers from inclusive classrooms with students with mild SEN.

First, the scale of inclusion of students with severe SEN is limited, despite the efforts expended by some scholars and parents (Kurth, Morningstar, & Kozleski, 2014; Ryndak et al., 2014). Thus, GETs do not have experience in teaching in classrooms with such students. Second, GETs often hold negative attitudes towards the inclusion of students with severe SEN (Avramidis & Norwich, 2002; de Boer, Pijl, & Minnaert, 2011). This negative attitude may translate into lower self-efficacy among teachers (Malinen, Savolainen, & Xu, 2012) and, consequently decrease their engagement in the classroom as well as adopting defensive adaptation strategies rather than seeking constructive methods for inclusive education. Third, students with severe SEN often require exclusively focused instructions, which absorbs teachers' attention (McDonnell, Thorson, & McQuivey, 2000). Therefore, GETs can have less time for students without SEN. Finally, students with severe SEN more often require intensive help in learning (Fuchs, Fuchs, & Vaughn, 2014), including one-to-one tutoring (McDonnell et al., 2006). For this reason, SETs in inclusive classrooms have less time and fewer opportunities to help other students in their studies, as well as fewer opportunities to use more effective co-teaching strategies, such as team teaching or station teaching (Scruggs, Mastropieri, & McDuffie, 2007).

1.3.5. Inclusive classrooms with and without students with EBD

The presence of students with EBD creates a challenge for inclusive education, because these students are among the most difficult categories of SEN (Avramidis, Bayliss, & Burden, 2000; Hornby & Evans, 2014). They make classroom management difficult, behave in a disruptive way and take-up a considerable amount of teachers' attention, even at the level of elementary education (Bruggink, Meijer, Goei, & Koot, 2014). Yet disruptive behavior is not the only problem these students may exhibit; they also have serious difficulties in assimilating academic knowledge (Bradley, Doolittle, & Bartolotta, 2008). These two problems create a difficult and absorbing task for general educators in inclusive classrooms when providing such students with the necessary support (Kauffman, Landrum, Mock, Sayeski, & Sayeski, 2005). Therefore, teachers often have negative attitudes toward including students with EBD in regular classrooms (Avramidis & Norwich, 2002). According to national reports, services for this group of students are the least developed compared to services for other groups of learners with SEN (Bradley et al., 2008; Wagner et al., 2006). Furthermore, there is a shortage of evidenced-based methods that would improve the behavior of these students (Bradley et al., 2008; Kauffman, Bruce, & Lloyd, 2012), as well as a shortage of special educators qualified to work with them (Bradley et al., 2008; Henderson, Klein, Gonzalez, & Bradley, 2005).

1.3.6. Educational stage

An educational stage may moderate the impact of inclusive education on achievement of students without SEN. It may be hypothesized that this effect is more positive in elementary than in middle and in high school classrooms. Students without SEN may improve in inclusive classrooms thanks to transforming teaching methods applied there. Additionally, they may have better access to individual help from teachers. This observation, however, fits poorly with high schools (Pearce & Forlin, 2005). GETs and SETs in high school rarely intensively cooperate to change teaching strategies in the classroom (de Vroey, Struyf, & Petry, 2016). High school classes, especially in Europe, are predominantly homogeneous, because of early tracking, thus GETs often use directed instruction for whole class teaching (King-Sears & Bowman-Kruhm, 2011; Murawski, 2006). Thus, it may be difficult for GETs to change their way of teaching, if necessary, because they have only a narrow repertoire of methods and teaching strategies. This stems from a much stronger emphasis on content knowledge than on instructional skills (Boe, Shinn, & Cook, 2007) in preparing high school teachers. Whereas GETs often use whole class teaching, SETs must concentrate on teaching students with SEN in small groups, sometimes even outside the classroom. They must be prepared to answer students' questions, because these students are often not able to follow the teacher or they cannot do the task by themselves (Weiss & Lloyd, 2002). Such roles of GETs and SETs do not foster planning classes, sharing experiences and preparing flexible whole class teaching strategies together.

Additionally, as high school GETs teach their subjects to many classes, they only have contact with particular students for a few hours each week. Hence, it becomes difficult for them to get to know each student well and to devote extra time to them outside of the classroom (Matzen, Ryndak, & Nakao, 2010; de Vroey et al., 2016). Moreover, many SETs are prepared for work with younger classes (Pearce & Forlin, 2005; Weiss & Lloyd, 2002). They do not know the high school curriculum well, and this is the reason why they are not able to effectively teach children without SEN.

1.3.7. Proportion of students with SEN in the classroom

The increase in the percentage of students with SEN in the classroom can cause a decrease of school achievement in their peers without SEN for three main reasons. First, students with SEN often display externalizing behaviors (i.e. Emerson, 2003), which may cause disruption during the lesson, distract the other students' attention while doing school tasks, or even force teachers to interrupt the planned lesson to re-order the classroom. Second, students with SEN need more instructions directed especially toward them, which can limit the time teachers could use for teaching other students in the classroom (Fuchs et al., 2014). Third, a large percentage of students with SEN in the classroom can cause burnout of GETs, and at the same time decrease their work engagement (Talmor et al., 2005).

A study of non-academic outcomes of students without SEN in inclusive classrooms (Gottfried, 2014) showed that while single students with SEN in the classroom did not negatively affect their peers, two to five such students in the classroom made peers' social functioning worse, while more than five students with SEN even strengthened this negative effect (Gottfried, 2014). This "threshold hypothesis" is yet to be replicated in the case of school achievement, but it seems plausible that similar effects may occur. Although a large study of school achievement of students without SEN in inclusive classrooms (Ruijs, Van der Veen, & Peetsma, 2010) did not find any relationship between the number of SEN students and their peers' achievement, these relationships may be non-linear as well.

1.3.8. Preparation of general education teachers

Preparing GETs to work in inclusive classrooms can play a key role in school achievement of students without SEN. Inclusive education often requires competencies other than traditional educational systems, from both GETs and from SETs (McCray, Butler, & Bettini, 2014). However, preparation of GETs seems to be crucial, because in most cases they make important instructional decisions in inclusive classrooms and have a dominant position in relation to special education teachers (Scruggs et al., 2007). However, systems of teachers' education rarely consider these new requirements. There are large differences both inside and between countries in access to courses on inclusive education for GETs (EADSNE, 2011). The effectiveness of GETs not only depends on finishing such a course, but also on the quality of the course. There are a few key aspects of these courses that impact how well teachers are prepared for work in inclusive classrooms. First, such preparation should not only consider methodical skills, but also beliefs and values of teachers (Forlin, 2010). Second, it should not be too focused on providing knowledge about disabilities, but rather concentrate on knowledge about the value of students' diversity in the classroom (Kozleski & Waitoller, 2010) and about effective strategies of teaching diversified groups (Jordan, Schwartz, & McGhie-Richmond, 2009). Such an approach implies development of other practical skills, i.e. instead of teaching teachers how to adjust tasks to individual educational needs they are taught universal design for learning, which can be valuable also for students without SEN (Theoharis & Causton-Theoharis, 2011).

1.3.9. Co-teaching models

Several co-teaching models may be used in inclusive classrooms (Scruggs et al., 2007). Indeed, scholars usually assume that all models can be effective, when they are used appropriately and adequately with the material taught and to address the needs of students (Hamilton-Jones & Moore, 2013). There is, however, no empirical evidence for effectiveness of different models for academic achievement of students with SEN as well as for students without SEN (Murawski & Swanson, 2001). Teachers usually use only one model of co-teaching, in which GET teaches the whole classroom, and SET supports students, who do not understand instruction, mostly students with SEN (Solis et al., 2012; Weiss & Lloyd, 2002). This model does not bring explicit profits for students without SEN, because it does not lead to implementing effective strategies and teaching methods or individualization, in the classroom (Murawski & Lochner, 2011; Solis et al., 2012). Applying this model does not mean that common lessons are being planned by both teachers. Alas, such cooperation is crucial for sharing knowledge and experiences, reciprocal skills enhancing, and preparing better instructional materials, i.e., consistent with a rule of universal design (Murawski & Lochner, 2011; Solis et al., 2012). In this model the role of SETs is to be nothing more than paid assistants (Murawski & Lochner, 2011). Other co-teaching models, such as station teaching, team teaching, alternative teaching or parallel teaching (Hamilton-Jones & Moore, 2013) can create more favorable conditions for effective learning by students without SEN. For example, alternative teaching in more homogeneous groups creates occasions for providing better adjusted instructions for weaker students, but without statements of disabilities (Solis et al., 2012); team teaching and parallel teaching may help to increase the number of interactions between students and teacher (Murawski & Lochner, 2011). Therefore, we hypothesize that using "one teach, one assist" model is less effective than using alternative models.

2. Method

2.1. Search strategies

To identify studies that examined the effects of inclusive education for students without SEN, we searched the following databases: EBSCO, PsychInfo, Academic Search Complete, Science Direct, ProQuest, and Questia, as well as reports published online and lists of references provided in the articles we found. We used the following search keywords in various configurations in both English and German: inclusive education, mainstream education, integrative education, effectiveness, co-teaching, achievement, school outcomes, students without disabilities, students without special educational needs, able students. We excluded publications in other languages, because of our lack of proficiency in these languages. We

acknowledge, however, that in several countries (Italy, France, Scandinavian countries) there are vast and diversified experiences in implementing inclusive education (Booth & Ainscow, 1998; Rix, Sheehy, Fletcher-Campbell, Crisp, & Harper, 2013). We sought relevant data in English-language and German-language journals that were devoted to special education and educational psychology on the websites of Elsevier, Sage, Wiley, and Taylor and Francis. This was guided by the subject matter of journals, by the titles of articles, or by keywords. This search was conducted manually based on tables of contents and abstracts for the years 1980–2013, or from the first issue in the case of younger journals. The English language journals included in this search were: *Exceptional Children, Journal of Special Education, Remedial and Special Education, International Journal of Inclusive Education,* and *European Journal of Special Needs Education*. The German language journals were: *Heilpädagogische Forschung, Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, Sonderpädagogische Förderung Heute,* and *Zeitschrift für Pädagogische Psychologie*. The search for German-language studies was supplemented by a survey at the library of the Faculty of Psychology and Pedagogy at the University of Munich and at the library of the University of Innsbruck. We decided to expand the analyses to also include studies found in reports that were available online as well as in doctoral dissertations and books. We included books because of the long tradition of publishing research results in this form in the German language area. We included doctoral dissertations and reports because many results on academic achievement of students without SEN are published in this form. However, because these works are not always peer-reviewed, we decided to assess the quality of studies included in our meta-analysis (see section: Assessment of Quality of Studies). We limited the search to those studies that were published between 1980 and 2015, but the oldest study we found dated back to 1987 (Wocken, 1987). This search yielded a total of 94 publications.

2.2. Criteria for including and excluding studies

To be included in our meta-analysis, studies had to meet the following criteria: be quantitative; academic achievement was measured using standardized tests; the measurement had to concern academic achievement in language, mathematics, science, biology, or a foreign language; participants were students in grades K–12; the study had to involve between-group comparisons, that is, cross-sectional (one measurement) or longitudinal (several measurements), or to be part of a one-group longitudinal design; the study had to report effect size or provide data necessary to compute it. We excluded studies in which academic achievement was measured with the use of grade point average as well as studies in which authors reported the results of research that had previously been presented in a different publication.

Our search identified 94 potentially relevant publications. They were considered for inclusion in the meta-analysis in two stages. In the first stage, we excluded 34 publications for six reasons (see Fig. 1). Data from the remaining 60 publications were subjected to further analyses. Then, a further 13 publications were excluded for two additional reasons (see Fig. 1).

The final dataset used in the meta-analysis included 47 publications: 18 from peer-reviewed journals, 1 (Wocken, 1987) chapter from an edited book, 3 studies (Feyerer, 1998; Haeberlin, Bless, Moser, & Klaghofer, 1999; Hinz, Katzenbach, Rauer, Schuck, Wocken, & Wudtke, 1998) obtained from research monographs, 3 research reports (Dyson et al., 2004; Friesen, Hickey, & Krauth, 2009; Herrmann, 2011), and the remaining 22 studies were doctoral dissertations defended in U.S. universities.

We extracted 138 effects from the 47 analyzed studies with a total of nearly 4 800 000 ($N = 4\,733\,985$) elementary, middle, and high school students (grades K-12). The studies were conducted between 1980 and 2012 in Europe (Austria, the Netherlands, Germany, Switzerland, the United Kingdom) and North America (Canada, United States). Detailed characteristics of the studies included in the meta-analysis are presented in Table 1.

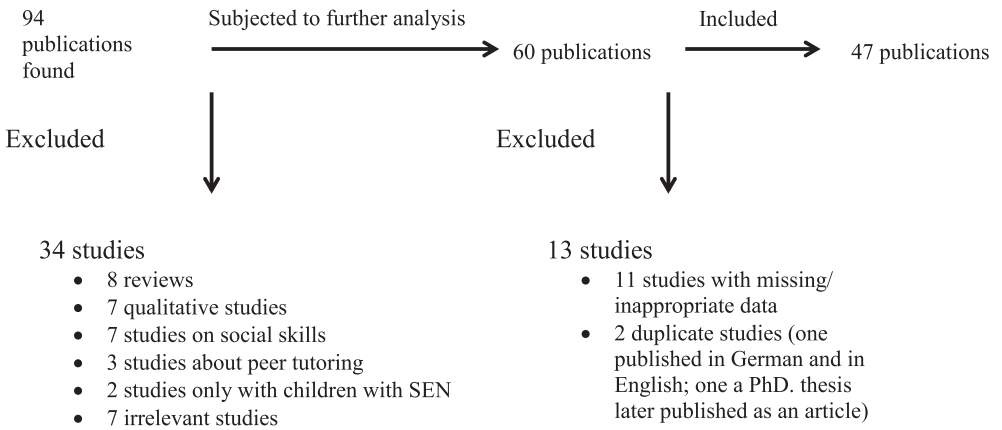

Fig. 1. Scheme of including and excluding studies for meta-analysis.

2.3. Assessment of the quality of studies

Prior to analyses, we performed an assessment of the quality of studies (Valentine, 2009; see Table 2). For this purpose, we applied the assessment criteria proposed by Dalemans, De Witte, Wade, and Van Den Heuvel (2008; see also Peltopuro, Ahonen, Kaartinen, Seppälä, & Närhi, 2014; Verdonschot, de Witte, Reichrath, Buntinx, & Curfs, 2009). This scheme consists of 15 items corresponding to 3 aspects of studies: informativity (I, 6 items), external validity (EV, 4 items), and internal validity (IV, 5 items). The first and the second authors of this article independently rated the studies. Inter-rater agreement was very good (Cohen's kappa = 0.72). After discussing the differences and reanalyzing the studies, the raters agreed upon the final assessments. Next, the quality of the studies, divided per publication type and study date, was compared using a nonparametric Mann–Whitney U test. Comparing peer-reviewed articles with the remaining publications, we found no significant differences in the global rating of study quality ($U = 270.5$, $p = 0.82$) or in its specific dimensions (I: $U = 270$, $p = 0.43$; EV: $U = 266.5$, $p = 0.90$; IV: $U = 250$, $p = 0.62$). The time when the studies were conducted (before or after 2000) also did not significantly differentiate their quality (overall: $U = 293$, $p = 0.19$; I: $U = 256$, $p = 0.14$; EV: $U = 300.5$, $p = 0.13$; IV: $U = 232.5$, $p = 0.72$).

2.4. Data coding

Two coders independently checked the data and any concerns were discussed and resolved by agreement. The first and the second authors of this paper independently coded the moderators. Each case was then discussed separately and any differences were resolved.[1]

When coding data, we supplied the following information: year of publication, country of study, research objective, field of knowledge tested (e.g., language, mathematics), study duration, number of measurements, study design, types and names of measures used, reliability of the measures, number of participants, participants' age and gender, form of education, number of teachers in the classroom, mean number of children in the class, number of children with SEN in the class, types and severity of disorders in the class, mean scores and standard deviations, results of the statistical tests used, information regarding whether data came from an experiment, and effect size. We entered these data whenever possible, however, some of the above information was missing in a few sources. Due to frequent cases of missing data we had to abandon the analysis of several potential moderators, such as class size, percentage of students with SEN in the classroom, co-teaching strategies, and instructional methods.

2.5. Moderators

We examined six moderators: country of study, implementation, educational team composition, level of SEN, type of SEN, and educational stage. For the analyses, country was binary-coded as U.S./Canada (coded 0) or European countries (coded 1). Thirty-six studies were from the United States, three from Canada and eight were from Europe.

Implementation was binary-coded as either implementation (coded 1) or regular school practice (coded 0). We defined implementation studies as those that concerned the evaluation of inclusive education implementation in a school, a group of schools, or a school district, carried out with support from university faculties or local educational authorities supported by inclusive education specialists. The implementations included training teachers, preparing new or modified educational programs or new methodological and organizational solutions, as well as organizing cycles of meetings for participants. We classified 13 studies as implementation studies and 34 as conducted in the regular education system.

The educational team composition was initially coded in two steps. At first, we coded a formal status of special education support, using two categories: (a) special education teacher, and (b) paraprofessional. Paraprofessionals were mentioned in only three studies (Gandhi, 2007; LaFever, 2012; McCartney, 2006), thus we removed them from an analysis of this moderator. Thereafter, we coded the number of teachers in the classroom using three categories: (a) only the GET, (b) an additional part-time SET, (c) a system of two full time teachers (GET and SET). In 26 studies, there were two teachers reported as working full time in the classroom. In nine studies, the second teacher in the classroom was reported as working part time. In the case of one study, there was only one teacher working in the classroom (Shinn, Powell-Smith, Good, & Baker, 1997). In one study, the authors compared two variants of inclusive education: one with two full-time teachers in the classroom and the other with the second teacher employed part-time (Hinz et al., 1998), so the study was excluded from the analyses of the role of this moderator. In nine studies, information about the number of teachers in the classroom was not given. Since in only one study class teachers received no support (Shinn et al., 1997), that study was excluded from the analyses of this moderator. Finally, the variable 'educational team composition' was coded as binary: full time general education teacher and special education teacher (coded 1) and part time special education teacher (coded 0).

Level of SEN was coded in two categories: severe SEN (coded 1) and mild SEN (coded 0); this is justified by a strong tradition in special education research (Cook, 2001). In coding this moderator, we used: (a) direct information provided in articles that indicated that the examined classrooms included students with severe or significant SEN or with mild SEN, and

[1] Divergent evaluations occurred only in the case of the implementation moderator. Three studies (Bean, 2006; Daniel & King, 1997; Shinn et al., 1997) were ambiguous. Following discussion, we decided that they did not meet the criteria for implementation studies.

Table 1
Studies included in the meta-analysis.

No.	Study/ Authors	Publication[1]	Year of study	Country	No. of participants (N)[2]	No. of effects	d	Moderators					
								Country of study[3]	Implementation[4]	Educational team composition[5]	Level of SEN[6]	Type of SEN[7]	Educational stage[8]
1.	Wocken (1987)	2 (chapter)	N/A (during one year)	Germany	234–496	3	0,33	1	1	1	1	1	0
2.	Affleck, Madge, Adams, and Lowenbraun (1988)	1	1980–1983	USA	78	1	0,02	0	1	0	0	1	0
3.	Dumke (1991)	1	1985–1988	Germany	141	12	−0,10	1	1	0	1	0	0
4.	Sharpe, York, and Knight (1994)	1	1988–1991	USA	143	4	−0,15	0	1	1	1	1	0
5.	Stevens and Slavin (1995)	1	N/A (2 years)	USA	107	6	0,01	0	1	1	0	0	0
6.	Willrodt and Claybrook (1995)	3	1995	USA	209–241	2	0,28	0	1	1	1	1	0
7.	Daniel and King (1997)	1	1994	USA	65	4	0,00	0	0	N/A	0	0	0
8.	Marcheski (1997)	3	1995–1996	USA	88	1	−0,24	0	0	1	1	0	1
9.	Shinn et al. (1997)	1	N/A (12 weeks)	USA	66	4	0,76	0	0	N/A	0	0	0
10.	Belmarez (1998)	3	1994–1997	USA	41	1	−0,15	0	0	1	0	0	1
11.	Feyerer (1998)	2 (book)	1996	Austria	651	3	−0,14	1	1	1	0	1	1
12.	Hinz, Katzenbach, Rauer, Schuck, Wocken, & Wudtke (1998)	2 (book)	1996	Germany	275	4	0,28	1	1	N/A	0	1	0
13.	Kettmann Klinger et al. (1998)	1	N/A (during one year)	USA	89	6	2,11	0	1	0	0	0	0
14.	Saint-Laurent et al. (1998)	1	1993–1994	Canada	441	3	0,10	0	1	0	0	1	0
15.	Haeberlin et al. (1999)	2 (book)	N/A (during one year)	Switzerland	350	2	0,13	1	0	0	0	0	0
16.	Erickson (2000)	3	1993–1994	USA	134	3	0,46	0	0	1	0	0	1
17.	Eichenholtz (2001)	3	1991–2000	USA	360	3	0,02	0	0	1	1	0	0
18.	Huber et al. (2001)	1	1992–1995	USA	477	4	0,21	0	1	N/A	0	1	0
19.	Cawley, Hayden, Code, & Baker-Kroczynski (2002)	1	N/A (during one year)	USA	98	2	−0,14	0	1	0	0	1	1
20.	Hanushek, Kain, and Rivkin (2002)	1	1993–1997	USA	767 763	1	0,01	0	0	N/A	1	1	0
21.	McDonnell et al. (2003)	1	N/A (during one year)	USA	545	2	0,03	0	1	1	1	0	0
22.	Cole et al. (2004)	1	1998–1999	USA	606	2	0,30	0	0	N/A	0	1	0

(*continued on next page*)

Table 1 (*continued*)

No.	Study/ Authors	Publication[1]	Year of study	Country	No. of participants (N)[2]	No. of effects	d	Moderators					
								Country of study[3]	Implementation[4]	Educational team composition[5]	Level of SEN[6]	Type of SEN[7]	Educational stage[8]
23.	Dyson et al. (2004)	4	2002	United Kingdom	1 796 354	8	0,00	1	0	N/A	1	1	0
24.	Eaton (2004)	3	2001 −2003	USA	317	1	0,17	0	0	0	0	1	0
25.	Brewton (2005)	3	2003 −2004	USA	1106	1	0,00	0	0	1	0	0	1
26.	Bean (2006)	3	2005 −2006	USA	206	1	−0,04	0	0	1	N/A	N/A	1
27.	McCartney (2006)	3	2003 −2004	USA	245–273	6	−0,07	0	0	1	1	0	0
28.	Castro (2007)	3	N/A (2 years)	USA	1023	3	0,04	0	0	1	0	0	0
29.	Demeris et al. (2007)	1	1997 −1998	Canada	~51 298	3	−0,02	0	0	N/A	1	1	0
30.	Gandhi (2007)	1	1998 −1999	USA	6293	1	0,10	0	0	1	1	1	0
31.	Neugebauer (2008)	3	2006 −2007	USA	4362–4381	2	−0,19	0	0	1	N/A	N/A	2
32.	Trejo (2008)	3	2006 −2007	USA	212–220	2	−0,12	0	0	0	1	1	0
33.	Friesen et al. (2009)	4	2002 −2004	Canada (British Columbia)	105 644 −107 603	2	−0,01	0	0	N/A	1	1	1
34.	Grimaldi (2009)	3	2006 −2007	USA	90	8	0,14	0	0	1	1	0	1
35.	Hartfield (2009)	3	2007 −2008	USA	462	1	−0,01	0	0	1	N/A	N/A	0
36.	Maultsby-Springer (2009)	3	2007 −2009	USA	182	2	−0,15	0	0	1	1	1	1
37.	Fletcher (2010)	1	1998 −2000	USA	11 373	2	−0,09	0	0	N/A	0	1	0
38.	Parker (2010)	3	2008 −2009	USA	103–122	2	−0,34	0	0	1	0	0	2
39.	Ruijs et al. (2010)	1	2004 −2005	The Netherlands	18 303–26 030	4	0,00	1	0	N/A	1	1	0
40.	Thompson (2010)	3	2008 −2010	USA	86	2	0,37	0	0	1	1	1	1
41.	Cuellar (2011)	3	2008 −2010	USA	30	2	1,38	0	0	1	0	0	2
42.	Sermier Dessemontet, Benoit, and Bless (2011)	1	2007 −2009	Switzerland	404	6	−0,02	1	0	0	0	0	0
43.	Herrmann (2011)	4	1999 −2011	USA	1 957 491	2	0,00	0	0	1	1	1	0
44.	Muscelli (2011)	3	2010 −2011	USA	124	1	0,17	0	0	1	N/A	N/A	2
45.	Trabucco (2011)	3	2009 −2010	USA	99	1	0,17	0	0	1	N/A	N/A	0
46.	Whisted (2011)	3	2008 −2009	USA	49	1	−0,16	0	0	1	0	0	2
47.	LaFever (2012)	3	N/A (during one year)	USA	140	1	1,04	0	0	N/A	0	0	2

Key:
[1]Publication: 1 – article, 2 – book/chapter, 3 – PhD. dissertation, 4 – other (report), paper officially unpublished.
[2] Number of participants can differ in the case of the analysis within the study.
[3] Country of the study: 0 – USA and Canada, 1 – other countries.
[4] Implementation: 0 – no, 1 – yes.
[5] Educational team composition: 0 – full time general education teacher and part-time special education teacher, 1 – full time general education teacher and special education teacher.
[6] Level of SEN: 0 – mild, 1 – severe.
[7] Type of SEN: 0 – classrooms without students with behavioral and emotional disorders, 1 – classrooms with students with behavioral and emotional disorders.
[8] Educational stage: 0 – elementary education, 1 – lower secondary education, 2 – upper secondary education.
N/A – data not available or not applicable.

(b) indirect data about specific types of students' SEN. In the latter case, we decided that there were students with severe SEN in the classroom if the authors reported the presence of students with multiple disabilities, moderate or severe intellectual disabilities, or Down syndrome or autism; this was in accordance with common research practice (Apler, 2003; Baker, Spooner, Ahlgrim-Delzell, Flowers, & Browder, 2010; Downing & Peckham-Hardin, 2007). Students with learning difficulties, EBD, communication disorders, sensory and motor disabilities, and mild intellectual disability were classified under the category of mild SEN. To avoid inequivalence, we considered the different language traditions in the categorization of SEN (Norwich, 2014). Given that in most inclusive classrooms there are students with various types of SEN, the "severe" category label was given to those classrooms in which there were students that we regarded as having severe SEN, regardless of the presence of other students with SEN. In 23 studies, only students with "mild SEN" attended the classrooms examined. In 19 studies, the information provided showed that there were students with more severe SEN in the classrooms. Five studies provided no information concerning the severity of disorders in students with SEN.

Type of SEN was binary-coded, with students with EBD (coded 1) distinguished from the remaining SEN types (coded 0). Coding was based on information provided in the description of the sample; as in the coding of level of SEN, the EBD label does not mean that a group included students exclusively with this type of SEN. Twenty of the studies noted no students with EBD in the examined classrooms, and in 22 studies the examined classrooms included students with these kinds of disorders. Five studies provided no information about students with emotional problems — these studies were excluded from the analyses of the role of this moderator.

Educational stage was classified in accordance with the commonly accepted division into elementary, middle, and high school, which corresponds to the division used by Eurostat and to the three tiers distinguished in this system: primary education, lower secondary education, and upper secondary education, as well as to the first, second, and third school levels in the International Standard Classification of Education (ISCED; UNESCO, 2012). Stage was coded based on the information given by the authors of the analyzed papers, which were then juxtaposed with the reported age of participants. In those sources that only provided the participants' age, classification was performed by juxtaposing the age with knowledge about the educational system in that given country and with ISCED. Thirty studies were conducted in grades K-5 (elementary education, coded 0), 11 concerned students attending grades 6–8 (lower secondary education, coded 1), and 6 pertained to students in grades 9-12 (upper secondary education, coded 2).

2.6. Statistical analyses

The data were analyzed in a three-level meta-analysis model (Cheung, 2014) due to the clustering of effects within studies, and in a more classic meta-analysis with random effects performed on data averaged to the level of individual studies (Schmidt & Hunter, 2014). The analyses were performed in R (R Development Core Team, 2013) using the metaSEM package (Cheung, 2014) and the metafor package (Viechtbauer, 2010). We analyzed the role of moderators with the use of a meta-analytic analog of ANOVA and the macros used for SPSS (Wilson, 2014).

The measure of effect size used in the meta-analysis was Cohen's d (Cohen, 1969): a standardized difference between achievement of students without SEN in classrooms with vs. without students with SEN. Effect size was computed based on descriptive statistics or statistical tests reported in the original studies. In one study (Ruijs et al., 2010), the exact effect size was not reported, but given that it was reported as non-significant on a very large sample ($N > 18.000$) we estimated this effect size as being exactly $d = 0$.

3. Results

The data comprised 143 effects clustered within 47 studies. We therefore began our analyses by estimating the mean effect size in a three-level meta-analytic model (see Cheung, 2014; for a more detailed discussion). This method makes it possible to accurately assess the variance assigned to the second level (within-study effects) and to the third level (the studies themselves). It turned out that the effect was positive and statistically significant, though weak: $d = 0.12$, $SE = 0.054$, 95% CI: 0.02, 0.23, $p = 0.02$. Second-level variance — pertaining to within-study effects — was significant, yet low ($s^2_2 = 0.0004$, $SE = 0.00019$, $p = 0.02$), and nearly all the variance was located between studies: $s^2_2 = 0.125$, $SE = 0.03$, $p < 0.001$, $I^2_2 = 0.35\%$, $I^2_3 = 99.6\%$. Due to the limited variability of within-study effects, we decided to perform all further analyses in a random effects model with results aggregated to the study level (Cheung, 2014).

The random effects model, estimated with the use of the restricted maximum likelihood method (REML, Viechtbauer, 2010), yielded an identical effect to that obtained in the three-level model, $d = 0.12$, $SE = 0.053$, 95% CI: 0.02, 0.23, $p = 0.02$, thus demonstrating a positive and significant, though weak, effect of the presence of students with SEN on the academic achievement of their peers without SEN. This effect was heterogeneous, $Q(df = 46) = 952.37$, $p < 0.001$, which points to the need to analyze potential moderators that may be responsible for the variability of these results. Prior to that, however, we examined the possible risk of publication bias.

3.1. Publication bias

We assessed the risk of publication bias in two steps. First, we compared the effects obtained in published and unpublished studies assuming that if publication bias were a problem in existing studies, then published studies should report significantly

Table 2
Studies' quality assessment.

	Study/Authors	Informativity							External validity					Internal validity						Overall
		a	b	c	d	e	f	sum	g	h	i	j	sum	k	l	m	n	o	sum	
1.	Wocken (1987)	+	+	+	+	+	+	6	+	+	−	−	2	+	+	+	+	+	5	13
2.	Affleck et al. (1988)	+	+	+	+	+	+	6	−	−	+	+	2	+	+	+	+	+	5	13
3.	Dumke (1991)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
4.	Sharpe et al. (1994)	+	+	+	+	+	+	6	−	+	+	+	3	+	+	+	+	+	5	14
5.	Stevens and Slavin (1995)	+	+	+	+	+	+	6	+	+	−	−	2	+	+	+	+	+	5	13
6.	Willrodt and Claybrook (1995)	+	−	+	+	+	+	5	−	+	−	+	2	+	+	−	+	+	4	11
7.	Daniel and King (1997)	+	+	+	+	+	+	6	−	−	−	+	1	+	+	+	+	+	5	12
8.	Marcheski (1997)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
9.	Shinn et al. (1997)	+	+	+	+	+	+	6	−	+	−	−	1	+	+	+	+	+	5	12
10.	Belmarez (1998)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
11.	Feyerer (1998)	+	+	+	+	+	+	6	+	+	+	+	4	+	+	+	+	+	5	15
12.	Hinz, Katzenbach, Rauer, Schuck, Wocken, & Wudtke (1998)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	12
13.	Kettmann Klinger et al. (1998)	+	+	+	+	+	+	6	−	+	−	−	1	+	+	+	+	+	5	12
14.	Saint-Laurent et al. (1998)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
15.	Haeberlin et al. (1999)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
16.	Erickson (2000)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
17.	Eichenholtz (2001)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
18.	Huber et al. (2001)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
19.	Cawley et al. (2002)	+	+	+	+	+	+	6	−	+	−	−	1	+	+	−	−	+	3	12
20.	Hanushek et al. (2002)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
21.	McDonnell et al. (2003)	+	+	+	+	+	+	6	−	+	+	−	2	+	+	+	+	+	5	13
22.	Cole et al. (2004)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
23.	Dyson et al. (2004)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
24.	Eaton (2004)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
25.	Brewton (2005)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
26.	Gandhi (2007)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
27.	Bean (2006)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
28.	McCartney (2006)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
29.	Castro (2007)	+	+	+	+	+	+	6	+	+	−	−	2	+	+	+	+	+	5	13
30.	Demeris et al. (2007)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
31.	Neugebauer (2008)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
32.	Trejo (2008)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
33.	Friesen et al. (2009)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
34.	Grimaldi (2009)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
35.	Hartfield (2009)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
36.	Maultsby-Springer (2009)	+	+	+	+	+	+	6	−	−	−	+	1	+	+	+	+	+	5	12
37.	Fletcher (2010)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	−	+	+	4	13
38.	Parker (2010)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
39.	Ruijs et al. (2010)	+	+	+	+	+	+	6	+	+	+	+	4	+	+	+	+	+	5	15
40.	Thompson (2010)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
41.	Cuellar (2011)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
42.	Sermier Dessemontet et al. (2011)	+	+	+	+	+	+	6	−	+	+	+	3	+	+	+	+	+	5	14
43.	Herrmann (2011)	+	+	+	+	+	+	6	+	+	−	+	3	+	+	+	+	+	5	14
44.	Muscelli (2011)	+	+	+	+	+	+	6	−	+	+	+	3	+	+	−	−	+	3	12
45.	Trabucco (2011)	+	+	+	+	+	+	6	−	+	+	+	3	+	+	+	+	+	5	14
46.	Whisted (2011)	+	+	+	+	+	+	6	−	+	−	+	2	+	+	+	+	+	5	13
47.	LaFever (2012)	+	+	+	+	+	+	6	−	+	−	−	1	+	+	+	+	+	5	12

Note: a – the purpose of the study is clearly described; b – the method of the data collection is clearly described; c – the main outcomes to be measured are clearly described in the introduction or methods section; d – the description of the characteristics of the population is sufficient; e – the response rate is >70% or the information of the no responders is sufficient; f – the main findings of the study are clearly described: simple outcome data should be reported for all major findings; g – the subjects asked to participate are representative for entire population from which they were recruited; h – the inclusion and exclusion criteria are described; i – the age range is specified; j – the study period is described; k – the data are prospectively collected; l – a comparison group is used and properly described; m – the measurement instrument(s) is/are described; n – the main outcome measures used are accurate (valid and reliable); o – age specific and gender specific outcomes are reported.

stronger effects. A comparison of results reported in published and unpublished studies revealed that the size of the effects obtained in them did not differ, $Q(df = 1) = 0.36$, $p = 0.55$. This indicated that we were not dealing with a situation of published studies reporting stronger effects. This finding suggests that publication bias did not significantly impact the obtained effect size.

The second step was the analysis of the relationships between effect sizes in particular studies and their statistical power, which we assessed by inspecting the funnel plot (Duval & Tweedie, 2000) and by means of statistical tests that could detect selective publishing. We supplemented the analysis with the trim-and-fill method (Duval & Tweedie, 2000), checking how the obtained effect would change if the results were distributed fully symmetrically around the mean.

Although the funnel plot (Fig. 2) suggested a slight asymmetry, particularly caused by the presence of two very strong effects ($d = 1.38$ and $d = 2.11$) in relatively low-powered studies (Cuellar, 2011; Kettmann Klinger, Vaughn, Hughes, Schumm,

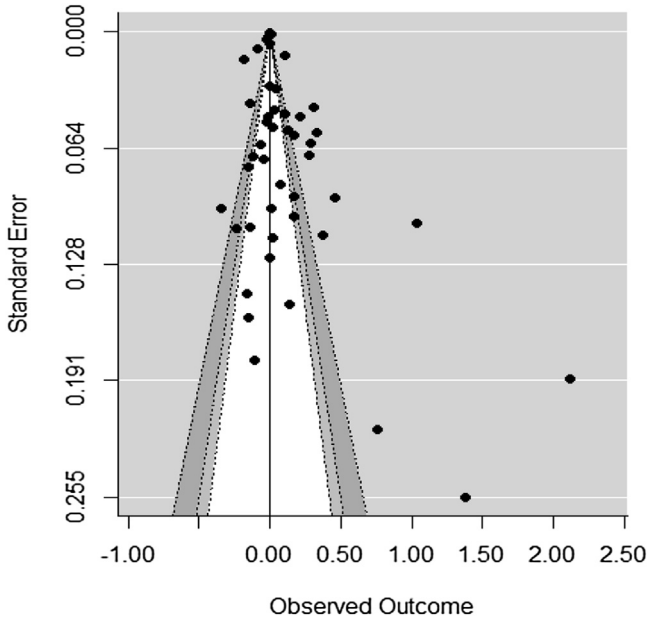

Fig. 2. Funnel plot examining the relationship between effect size (d) and standard error in studies.

& Elbaum, 1998), the correlation between effect size and standard error was not statistically significant (Kendall's tau = 0.11, $p = 0.27$). Although the two clearly outlying effects might suggest publication bias, two issues effectively make that risk small. First, only one of them (Kettmann Klinger et al., 1998) came from a published study, whereas the other was reported in an unpublished doctoral dissertation. Second, one of these effects (Kettmann Klinger et al., 1998) came from an implementation (intervention study), and studies of that kind are conducted on smaller samples, which translates into higher standard error.

The trim-and-fill method, which forces a symmetrical distribution of effects around the mean, did not indicate that additional studies with negative effects are required to obtain full symmetry. This confirms that our findings are robust for publication bias. At the same time, it must be remembered that the basic assumption underlying the trim-and-fill method — namely, the assumption of perfect symmetry in the distribution of effects around the mean — is not very realistic (Peters, Sutton, Jones, Abrams, & Rushton, 2007).

Finally, we checked the sensitivity of the obtained estimation with the presence of outlying effects. For this purpose, we estimated the overall effect again, first excluding the study that yielded the strongest effect ($d = 2.11$), then excluding the second strongest effect ($d = 1.38$), and then excluding four studies characterized by very high power due to vast samples ($N = 106\,624$; 767 763; 1 796 354, and 1 957 491 respectively). The summary of these analyses is presented in Table 3.

Although excluding the strongest effects resulted in a considerable (though statistically non-significant) reduction in the strength of the observed relationship, it remained positive and different from zero. Nevertheless, considerable heterogeneity of the overall effect size makes it necessary to perform further analyses to test the significance of moderators.

3.2. Moderator analysis[2]

The role of each moderator was tested in a meta-analytic ANOVA analog design with the use of Wilson's (2014) set of macros. The effects obtained for different levels of each moderator are presented in Table 4.

The first of the moderators analyzed was the country in which the study was conducted. We expected that U.S. and Canadian studies would yield significantly stronger effects than the analogous European studies. This expectation was largely confirmed: although the differences between effects were not statistically significant, $Q(df = 1) = 0.15$, $p = 0.70$, only the effect

[2] Apart from testing the possible significance of the moderators mentioned below, we also checked if the year when the study was conducted differentiated the obtained effects. Meta-regression analysis with study year predicting effect size did not reveal differences that depended on the study year ($B = -0.0067$, $SE = 0.0066$, $\beta = -0.12$, $p = 0.31$). Given that inclusive practices may have changed under the influence of the 1994 Salamanca Statement (UNESCO, 1994), we also checked whether studies conducted up to 1994 yielded different effects from those published after 1994. Here we also found no significant differences between effects, $Q = 0.03$, $df = 1$, $p = 0.81$.

Table 3
Robustness check of effect size with outliers and high-powered studies excluded.

Moderator	*k*	*d*	*SE*	95% CI LB	95% CI UB
Overall effect	47	0.12*	0.05	0.02	0.23
One outlier excluded	46	0.08*	0.04	0.01	0.16
Two outliers excluded	45	0.07*	0.002	0.002	0.13
Two outliers and four most powered studies excluded	41	0.08*	0.04	0.002	0.15

Note. * $p < 0.05$; k = number of studies; d = Cohen's d; SE = standard error; 95% CI LB UB = 95% confidence intervals: lower and upper bounds.

Table 4
Moderator analysis: ANOVA analog results.

Moderator	*k*	*d*	*SE*	95% CI LB	95% CI UB
Country of the Study					
USA/Canada	39	0.14*	0.07	0.02	0.27
Other	8	0.06	0.06	−0.06	0.18
Implementation					
Regular practice	34	0.06	0.05	−0.03	0.15
Intervention	13	0.18*	0.08	0.02	0.35
Educational Team Composition					
Full time general education teacher and part-time special education teacher	26	0.05	0.07	−0.09	0.18
Full time general education teacher and special education teacher	9	0.19	0.11	−0.03	0.41
Level of SEN					
Mild	23	0.19**	0.07	0.07	0.32
Severe	19	0.02	0.07	−0.11	0.16
Behavioral problems					
No	20	0.19**	0.08	0.04	0.33
Yes	22	0.06	0.06	−0.07	0.19
Educational stage					
Elementary education	30	0.11*	0.05	0.02	0.21
Lower secondary education	11	0.004	0.09	−0.18	0.19
Upper secondary education	6	0.20	0.13	−0.07	0.46

Note. k = number of studies; d = Cohen's d; SE = standard error; 95% CI LB = lower bound of 95% confidence intervals; 95% CI UB = upper bound of 95% confidence intervals.
*$p < 0.05$; **$p < 0.01$.

observed in the U.S. and Canada ($d = 0.14$) was statistically significant ($p = 0.03$), whereas the effect obtained in the remaining countries ($d = 0.06$) did not differ significantly from zero ($p = 0.33$).[3]

The next moderator included in our analyses was the manner of implementation. We compared the effects found in research on the implementation of experimental interventions with research describing school reality in which no changes were introduced. Here the effects did not differ one from another, $Q(df = 1) = 1.61$, $p = 0.20$; although the effect was non-significant in studies that described regular practice, $d = 0.06$, $p = 0.19$, it was significant in cases of intervention, $d = 0.18$, $p = 0.01$.

The effect observed in studies in which there were two simultaneously functioning full-time teachers (GET and SET) employed was not significant ($d = 0.05$, $p = 0.50$); this was similar to the effect observed in those studies in which the special education teacher was part-time ($d = 0.19$, $p = 0.09$). These effects did not differ significantly one from another, $Q(df = 1) = 1.26$, $p = 0.26$.

Differences were marginal between effects obtained in classrooms attended by students with severe SEN and in those attended by students with mild SEN, $Q(df = 1) = 3.22$, $p = 0.07$. As expected, we found positive effects with the presence of students with mild SEN on the achievement of their classmates without SEN ($d = 0.19$, $p = 0.003$), whereas no such effect was present in classrooms that included students with severe SEN ($d = 0.0.02$, $p = 0.74$).

We did not find significantly different effects, $Q(df = 1) = 1.48$, $p = 0.22$, between classrooms attended by students with EBD ($d = 0.06$, $p = 0.19$), and classrooms without such students, although a positive effect was found in classrooms without students with EBD ($d = 0.19$, $p = 0.01$).

[3] As in England the legislator encourages implementing inclusive education understood as actions of excising barriers in learning for all students, which makes England less similar to other European countries, we decided to conduct an additional analysis excluding one study from this country from the group of European countries. After this exclusion, the obtained effect size for continental European countries was $d = 0.08$, 95% *CI*: −0.14, 0.30; $p = 0.48$. The difference between this effect and effects obtained in the USA/Canada was not significant, $Q(df = 1) = 0.07$; $p = 0.79$. We are grateful to an anonymous reviewer for raising this to our attention.

The last moderator we analyzed was educational stage. The observed effects also did not differ significantly, $Q(df = 1) = 1.75$, $p = 0.42$. The effect observed in elementary education was weak but statistically significant ($d = 0.11$, $p = 0.03$), whereas non-significant effects were observed in upper secondary ($d = 0.20$, $p = 0.14$) and lower secondary education ($d = 0.004$, $p = 0.97$).

4. Discussion

This meta-analysis, exploring studies on the academic achievement of students without SEN in inclusive education, addresses a problem of considerable social and theoretical importance. In this section, we discuss our results in three independent contexts. First, we attempt to show their significance for the pragmatic, and legal legitimization of inclusive education. Second, we discuss these results in the context of the concept of inclusive education, understood as school for all and for everyone. Third, we address the hypotheses we have formulated about the role of the moderators that are theoretically and practically important but are not directly related to the legitimization of inclusive education or to the idea of school for all.

4.1. Academic achievement of students without SEN and the legitimization of inclusive education

The main finding of this meta-analysis shows that not only students with SEN may benefit from this form of education, which has been demonstrated in all the meta-analyses performed so far (Carlberg & Kavale, 1980; Oh-Young & Filler, 2015; Wang & Baker, 1985–1986) and in many massive studies addressing this issue (Cosier, Causton-Theoharis, & Theoharis, 2013; Rea, McLaughlin, & Walther-Thomas, 2002; Szumski & Karwowski, 2014), but also that inclusive education may be beneficial for students without SEN. This result may be important to educational policy-makers responsible for decisions about the promotion of inclusion, but also to parents of children without SEN. Although parents from this group usually have positive or neutral attitudes towards inclusion (de Boer, Pijl, & Minnaert, 2010), their attention is focused on the benefits for the social development of their children. Parents often fail to see the academic benefits and even have problems understanding why such benefits could occur in the first place (Peck, Staub, Gallucci, & Schwartz, 2004). This meta-analysis suggests that such academic benefits are plausible.

Even more importantly, the main effect of this meta-analysis supplements and supports argumentation in favor of promoting inclusion. The arguments used so far have been strongly focused on students with SEN and legal arguments have dominated discussions (Dyson, 1999; Miles & Singal, 2010). Only to a very small extent has the current way of justifying inclusive education considered students without SEN and their right to individual achievement (Artiles, Harris-Murri, et al., 2006). We perceive the evidence that inclusive education does not infringe upon the rights of the majority of students as especially important. We tested six moderators that in total could take 13 values that correspond to various conditions of inclusive education implementation. In none of these conditions did we find an effect (d) with a significantly negative value, which — if found — might have suggested a negative impact of inclusion on the academic achievement of students without SEN. All these effects were, at worst, neutral, and many of them were positive indeed (see Table 4). Even the presence of students with EBD in the classroom, usually hypothesized to have negative effects on inclusive education (Fletcher, 2010), did not lead to a significant negative effect.

4.2. Academic achievement of students without SEN and inclusive education concept

The effects we obtained — both the main effect and the results of moderator analysis — consistently support the concept of inclusive education, understood as effective school for all (Ainscow et al., 2012). For over a decade now, inclusive education has no longer been treated as a concept of special education but as a more radical concept of educational system transformation (Opertti et al., 2014) and building a school that ensures access and high achievement for all children (Ainscow et al., 2012). Although the results of our analyses are consistent with this assumption, they do not suggest that inclusive education alone can bring radical improvement in the quality of education. The effect size of $d = 0.12$ shows that the achievement of students without SEN in inclusive classrooms is only slightly better (the difference is equivalent to 2 points on a scale with a mean of 100 and a standard deviation of 15 points) than the achievement of their peers in non-inclusive classrooms. Nevertheless, this result must not be disregarded as the effectiveness of few interventions and strategies in education is high, particularly if they concern the school level or the school system rather than instruction strategies (Hattie, 2009; Hattie & Yates, 2014). For example, the effect size of factors such as educational expenditure ($d = 0.23$), class size ($d = 0.21$), ability grouping ($d = 0.12$), or within-class grouping ($d = 0.16$) is similar (Hattie, 2009) even though there is a firm and deeply rooted belief in their influence on the quality of education in schools. The effect of the inclusive classroom is also only slightly lower than the effect of introducing charter schools ($d = 0.20$) (Hattie, 2009), one of the key projects in the neo-conservative reform of school systems in the 1990s (Finn, Manno, & Vanourek, 2000).

In the context of the debate on inclusive education as a concept of school transformation, the effects of the moderators we tested are slightly more ambiguous. What supports this concept is the comparison of effect sizes obtained in studies conducted in the U.S. and Canada with those obtained in European countries. Although these effects did not differ significantly, a

statistically significant effect was obtained only in the U.S. and Canada ($d = 0.14$). We explain these results by the stronger experience in the promotion of inclusive education in the U.S. and Canada as compared to Europe (Armstrong & Barton, 2008; McLeskey, Landers, Williamson, & Hoppey, 2012; Powell, 2009), as well as the different way of understanding inclusive education. In the U.S., inclusive education is treated as a process of transforming general education, whereas in most European countries it is understood as a form of instruction for students with SEN (Norwich, 2008). Most of the European studies included in the meta-analysis originate from countries which have segregated school systems. They not only send most students with SEN to special schools (two-track systems are used in Switzerland, Germany, The Netherlands), but also send students without SEN to different types of schools early (European Commission, 2014; Norwich, 2008b). These systems, thus, seem to be based, often implicitly, on three premises: (a) academic achievement is more important than other values, like egalitarianism or community; (b) students should learn in homogeneous groups; (c) decisions are made on the basis of the medical model of disability (Pfahl & Powell, 2011; Powell, 2009). These premises conflict with inclusion understood as a principal approach to education (Ainscow et al., 2006; McLeskey & Waldron, 2006). Therefore, the effect of this moderator may be understood as showing that in European countries there is room for improving the quality of inclusion and for using this concept as a way to transform the school system.

Although the effect sizes obtained in implementation studies and regular school practices did not differ significantly, we obtained a positive effect only in intervention programs. Many solutions recommended in the literature were used in the implementations examined in the meta-analysis (Farrell et al., 2007; McLeskey, Waldron, Spooner, & Algozzine, 2014; McLeskey et al., 2014). These included the improvement of teacher training by incorporating in-service training; strong leadership of district and school authorities; planning, monitoring, and modifying the implementation; cooperation of the people involved in the project; cooperative learning; group instruction; clear expectations from students; and other recommended teaching strategies (Cawley, Hayden, Code, & Baker-Kroczynski, 2002; Dumke, 1991; Huber, Rosenfeld, & Fiorello, 2001; Kettmann Klinger et al., 1998; Saint-Laurent et al., 1998). This outcome is therefore consistent with the idea of inclusive education and demonstrates that changing a regular school into an inclusive one may improve the academic achievement of students without SEN. The aim of intervention programs is not only to give support to students with SEN. It is also important that all school personnel are engaged in inclusion implementation, and all people have in mind that they work to improve the functioning of a school (i.e. Kettman Klinger et al., 1998; Stevens & Slavin, 1995).

4.3. Support from special education teachers and the academic achievement of students without SEN

One of the moderators we tested was the engagement of special education teachers in inclusive classes. This problem is often discussed in the context of the achievement of students with SEN (Bottge et al., 2015; Zigmond, 2006), but much less attention is devoted to the role of SETs in the achievement of students without SEN in inclusive classrooms. Contrary to expectations, we did not find any significant difference in the effect size obtained in classrooms with GET and SET working full time and/or part-time. It should be stressed that we were comparing only forms that differed in the amount of time that SETs were engaged in the classroom. Since students were taught by one teacher in only one study included in our meta-analysis (Shinn et al., 1997), we were unable to examine how the academic achievement of students without SEN is affected by the total lack of a special education teacher in the classroom. Apart from that, it is worth noting that we analyzed achievement in academic subjects (primarily language, mathematics, and science), and during lessons in these subjects, SETs were always present. A study of social competence or other skills might yield different results. Perhaps it is also the case that schools are flexible and able to adjust the extent of support by SETs per the conditions in their classrooms, the number of students in classes, the number of students with SEN and the type of those needs, and the stage of education. We were unable to examine these suppositions due to the lack of relevant data, but this issue requires further research. This is because little is known about the impact of co-teaching on the academic achievement of students without SEN in inclusive classrooms, and, furthermore, the results of studies on the significance of this factor for the achievement of students with SEN are ambiguous. Co-teaching alone does not guarantee that more modern teaching strategies, beneficial for students both with and without SEN, will be introduced in the classroom. In numerous classrooms, the role of SETs is limited to supporting learners with SEN, which may not benefit learners without SEN. However, we did not control for models of co-teaching due to the lack of data. A recent study by Bottge et al. (2015) clearly suggests that an increase in the participation of SETs in instructional activities improves the academic achievement of students with SEN. This result is consistent with the thesis advanced by Zigmond (2006) that SETs should instruct students with SEN in learning strategies, since GETs, especially in higher grades, focus on providing new knowledge. At the same time, the research shows that SETs in inclusive classrooms not only fail to do this but, in fact, engage in direct instruction to a very limited extent (Zigmond, 2006).

4.4. Students' type of SEN and the academic achievement of their classmates without SEN

In our analysis, we tested the role of two moderators in the form of students' types of SEN. Contrary to expectations, we did not find differences in effect size when comparing classrooms with students with severe SEN and those with EBD to classrooms without such students, although in the former case the difference shows a statistical tendency. At the same time, we observed a statistically significant, positive, yet weak effect solely in classrooms in which there were no children with severe

SEN or EBD. Although, as mentioned in the introduction, there are many arguments in favor of the hypothesis postulating the negative influence of these two groups of learners on the quality of inclusive education, especially on the school achievement of students without SEN, previous studies yielded ambiguous findings. Some indicated the negative influence of children with EBD on the achievement of their classmates without SEN (Cawley et al., 2002; Fletcher, 2010; Friesen et al., 2009), yet no such influence was noted in other studies (Hanushek, Kain, & Rivkins, 2002). Furthermore, some studies demonstrated the lack of a negative influence of students with severe SEN on the achievement of their peers who had no SEN (Downing & Peckham-Hardin, 2007; Hollowood, Salisbury, Rainforth, & Palombaro, 1994; Katz, Mirenda, & Auerbach, 2002). These studies, however, do not provide strong arguments because they did not include a measurement of academic achievement but only the opinions of parents, teachers, and paraeducators (Downing & Peckham-Hardin, 2007), or data concerning teachers' use of time for instruction and students' engagement time (Hollowood et al., 1994). In conclusion, our meta-analysis shows that on average the presence of learners with EBD and severe SEN in a classroom does not negatively influence the achievement of their peers without SEN. Still, neither does it influence their achievement positively, nor is it conducive to the full use of the potential of inclusive education. Further studies are necessary to determine if this is the case regardless of the way inclusive education is organized.

4.5. Educational stage and students' academic achievement

The last moderator we considered was educational stage. Drawing on previous studies that suggest that the organization of inclusive education on higher school levels is more difficult (Pearce & Forlin, 2005) and less acceptable to teachers (Avramidis & Norwich, 2002) than is the case at lower levels, we expected, that the academic achievement of elementary school students would be significantly higher than that of students attending middle and high schools. Although we found no significant differences between effects from the three educational stages, an expected, positive effect was observed only in elementary education ($d = 0.11$). At the same time, we did not find a negative effect in lower and upper secondary education, quite the opposite — the effect in upper secondary education was visible ($d = 0.20$), even if not statistically significant ($p = 0.14$). Given the many potential difficulties in the implementation of inclusive education in high schools (Pearce & Forlin, 2005), it is introduced much less frequently at this stage. The effectiveness of inclusive education in high schools certainly requires future research as, to date, studies concerning inclusive education at lower educational stages have predominated.

5. Limitations and future studies

Although the present study contributes to the growing literature on inclusive education, it is not without limitations. First, the studies included in our meta-analysis had various designs. Although a longitudinal design with a control group is the strongest and most appropriate one to answer our questions, only a few studies were carried out in this manner. In several studies, authors used a cross-sectional design or a longitudinal one without a control group. Second, we did not analyze how specific subjects of study moderate the obtained effect, even if some studies indicated that it may be stronger in reading and language than in math (Murawski & Swanson, 2001). This issue requires further study. Third, the results we meta-analyzed came from the U.S., Western Europe, and Canada. We found no studies from Eastern Europe, Africa, Asia, or South America, even though in many of those regions inclusive education has been gaining ground (Alkhateeb, Hadidi, & Alkhateeb, 2016; Mu, 2015). Nevertheless, as our analysis has shown, the country of study does have some impact on the achievement of students without SEN in inclusive classrooms. For this reason, great caution is necessary when extrapolating the results, we obtained for inclusive education to countries in which the educational context is different from that in the U.S., Western Europe, or Canada. Fourth, our meta-analysis revealed high heterogeneity of the main effect, thus pointing to the need to analyze the role of moderators. We managed to include six moderators but the lack of data made it impossible to test the role of many other theoretically justified factors, such as class size, proportions between students with and without SEN in the classroom, teachers' preparation, or co-teaching models. This research problem deserves special attention in future studies. Fifth, we were unable to examine the role of inclusive education for the academic achievement of students with different levels of ability but without SEN. This problem is discussed quite regularly (Huber et al., 2001; Ruijs & Peetsma, 2009; Sermier Dessemontet & Bless, 2013) but has not yet been resolved. Huber et al. (2001) demonstrated that learners with low achievement benefit from inclusive education to a greater degree than those with higher achievement, but this result was not replicated in a recent study (Sermier Dessemontet & Bless, 2013). Sixth, being limited to students' school achievement, our meta-analysis did not comprehensively examine all the assumptions behind the transformative version of inclusive education (Opertti et al., 2014). We excluded students' social competence and emotional development in inclusive classrooms from our research, and we also excluded social relations, which are difficult to operationalize. We also could not analyze the culture of schools and its relations with local societies, which are often emphasized in the literature (Ainscow & Miles, 2008; Artiles, Kozleski, et al., 2006).

6. Conclusion

The main finding of this meta-analysis may be summarized as: attending inclusive classrooms is positively, though weakly, associated with the academic achievement of students without SEN. This result provides advocates of inclusive education

with important arguments in the ongoing debate about promoting this concept but, keeping in mind the size of the effect we obtained ($d = 0.12$), this should not be overestimated.

Acknowledgement

This research was supported in part by Contract No. UMO-2012/07/HS6/01442 from the National Science Centre Poland.

References

*denotes studies included in the meta-analysis

Adair, J. G., Sharpe, D., & Huynh, G.-L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research, 59*, 215–228.

* Affleck, J. Q., Madge, S., Adams, A., & Lowenbraun, S. (1988). Integrated classroom versus resource model: Academic viability and effectiveness. *Exceptional Children, 54*, 339–348.

Ainscow, M., Booth, T., & Dyson, A. (2006). *Improving schools, developing inclusion*. London and New York: Routledge.

Ainscow, M., & Cesar, M. (2006). Inclusive education ten years after Salamanca: Setting the agenda. *European Journal of Psychology of Education, 21*, 231.

Ainscow, M., Dyson, A., Goldrick, S., & West, M. (2012). Making schools effective for all: Rethinking the task. *School Leadership & Management: Formerly School Organisation, 32*, 197–213.

Ainscow, M., & Miles, S. (2008). Making education for all inclusive: Where next? *Prospects, 38*, 15–34.

Alkhateeb, J. M., Hadidi, M. S., & Alkhateeb, A. J. (2016). Inclusion of children with developmental disabilities in arab countries: A review of the research literature from 1990 to 2014. *Research in Developmental Disabilities, 49–50*, 60–75. http://dx.doi.org/10.1016/j.ridd.2015.11.005.

Apler, S. (2003). Students with moderate and severe disabilities: Definitions and descriptive characteristics. In D. L. Ryndak, & S. Apler (Eds.), *Curriculum and instruction for students with significant disabilities in inclusive settings* (pp. 1–12). Boston: Allyn & Bacon.

Armstrong, F., & Barton, L. (2008). Policy, experience and change and challenge of inclusive education: The case of England. In L. Barton, & F. Armstrong (Eds.), *Policy, experience and change: Cross-cultural reflections on inclusive education* (pp. 5–18). Dordrecht: Springer Netherlands.

Artiles, A. J., Harris-Murri, N., & Rostenberg, D. (2006). Inclusion as social justice: Critical notes on discourses, assumptions, and the road ahead. *Theory Into Practice, 45*, 260–268.

Artiles, A. J., Kozleski, E., Dorn, S., & Christstensen, C. (2006). Learning in inclusive education research: Re-mediating theory and methods with a transformative agenda. *Review of Research in Education, 30*, 65–108.

Avramidis, E., Bayliss, P., & Burden, R. (2000). A survey into mainstream teachers' attitudes towards the inclusion of children with special educational needs in the ordinary school in one local education authority. *Educational Psychology, 20*, 191–211.

Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education, 17*, 129–147.

Baker, J. N., Spooner, F., Ahlgrim-Delzell, L., Flowers, C., & Browder, D. M. (2010). A measure of emergent literacy for students with severe developmental disabilities. *Psychology in the Schools, 47*, 501–513.

* Bean, N. E. (2006). *Effects of co-teaching in the inclusive middle school classroom on student achievement, student attitudes, and teacher attitudes*. Unpublished doctoral dissertation. New Britain, CT: Central Connecticut State University.

* Belmarez, B. L. (1998). *The relationship between co-teaching and the mathematic achievement of groups of seventh-grade students with and without learning disabilities*. Unpublished doctoral dissertation. Kingsville, TX: Texas A & M University.

Boe, E., Shinn, S., & Cook, L. H. (2007). Does teacher preparation matter for beginning teachers in either special or inclusive education? *Journal of Special Education, 41*, 158–170.

Booth, T., & Ainscow, M. (Eds.). (1998). *From them to us. A international study of inclusion in education*. London and New York: Routledge.

Booth, T., & Ainscow, M. (2002). *Index for inclusion: Developing learning and participation in schools*. Bristol: Centre for Studies on Inclusive Education.

Bottge, B. A., Toland, M. D., Gassaway, L., Butler, M., Choo, S., Griffen, A. K., et al. (2015). Impact on enhanced anchored instruction in inclusive math classrooms. *Exceptional Children, 81*, 158–175.

Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education, 17*, 4–23.

* Brewton, S. (2005). *The effects of inclusion on mathematics achievement of general education students in middle school*. Unpublished doctoral dissertation. South Orange, NJ: Seton Hall University.

Bruggink, M., Meijer, W., Goei, S. L., & Koot, H. M. (2014). Teachers' perception of additional support needs of students in mainstream primary education. *Learning and Individual Differences, 30*, 163–169.

Carlberg, C., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education, 14*, 295–309.

* Castro, V. E. (2007). *The effect of co-teaching on academic achievement of K-2 students with and without disabilities in inclusive and non-inclusive classrooms*. Unpublished doctoral dissertation. New York, NY: Fordham University.

Causton, J., & Theoharis, G. (2014). How do schools become effective and inclusive? In J. McLeskey, N. L. Waldron, F. Spooner, & B. Algozzine (Eds.), *Handbook of effective inclusive schools* (pp. 30–42). New York, London: Routledge.

* Cawley, J., Hayden, S., Code, E., & Baker-Kroczynski, S. (2002). Including students with disabilities into the general education science classroom. *Exceptional Children, 68*, 423–435.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 2*, 211–229.

Cohen, J. (1969). *Statistical power analysis for behavioral sciences*. San Diego, CA: Academic Press.

* Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental Retardation, 42*, 136–144.

Connor, D. J., & Ferri, B. A. (2007). The conflict within: Resistance to inclusion and other paradoxes in special education. *Disability & Society, 22*, 63–77.

Cook, B. G. (2001). A comparison of teachers' attitudes toward their included students with mild and severe disabilities. *Journal of Special Education, 34*, 203–213.

Cosier, M., Causton-Theoharis, J., & Theoharis, G. (2013). Does access matter? Time in general education achievement for students with disabilities. *Remedial and Special Education, 34*, 323–332.

* Cuellar, K. (2011). *The effect of collaborative teaching on the general education student population: A case study*. Unpublished doctoral dissertation. Houston, TX: University of Houston.

Cusumano, D. L., Algozzine, K., & Algozzine, B. (2014). Multi-Tiered System and supports for effective inclusion in elementary schools. In J. McLeskey, N. L. Waldron, F. Spooner, & B. Algozzine (Eds.), *Handbook of effective inclusive schools* (pp. 197–228). New York, London: Routledge.

Dalemans, R. J. P., De Witte, L. P., Wade, D. T., & Van Den Heuvel, W. J. A. (2008). A description of working-age persons with aphasia: A review of the literature. *Aphasiology, 22*, 1071–1091.

* Daniel, L. G., & King, D. A. (1997). Impact of inclusion education on academic achievement, student behavior and self-esteem, and parental attitudes. *The Journal of Educational Research, 91*, 67–80.

de Boer, A., Pijl, S. J., & Minnaert, A. (2010). Attitudes of parents towards inclusive education: A review of the literature. *European Journal of Special Needs Education, 25*, 165–181.
de Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education, 15*, 331–353.
* Demeris, H., Childs, R. A., & Jordan, A. (2007). The influence of students with special needs included in grade-3 classrooms on the large-scale achievement scores of students without special needs. *Canadian Journal of Education, 30*, 609–927.
Downing, J. E., & Peckham-Hardin, K. D. (2007). Inclusive education: What makes it a good education for students with moderate to severe disabilities? *Research & Practice for Persons With Severe Disabilities, 32*, 16–30.
* Dumke, D. (1991). Schulleistungen nichtbehinderter Schüler in Inegrationsklassen. *Zeitschrift für Pädagogische Psychologie, 5*, 33–42.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.
Dyson, A. (1999). Inclusion and inclusions: Theories and discourses in inclusive education. In H. Daniels, & P. Garner (Eds.), *Inclusive education* (pp. 36–53). London: Kogan Page.
* Dyson, A., Farrell, P., Polat, F., Hutcheson, G., & Gallannaugh, F. (2004). *Inclusion and pupil achievement*. Research Report. Newcastle, DE: University of Newcastle.
EADSNE (European Agency for Development in Special Needs Education). (2011). Teacher education for inclusion across Europe. *Challenges and opportunities*. Retrieved from https://www.european-agency.org/sites/default/files/te4i-challenges-and-opportunities_TE4I-Synthesis-Report-EN.pdf.
* Eaton, D. (2004). *A comparative study of reading achievement among general education students in elementary third-grade inclusive and traditional classrooms in selected northern California schools*. Unpublished doctoral dissertation. La Verne, CA: University of La Verne.
Education for All Handicapped Children Act. (1975). Public law 94–142. Retrieved from www.gpo.gow/fdsys/pkg/STATUE-89/pdf/STATUE-89-Pg773.pdf.
* Eichenholtz, D. A. (2001). *The effects of the inclusion/main streaming of special education students on the academic performance and social tolerance of general education students*. Unpublished doctoral dissertation. Jamaica, NY: St. John's University.
Emanuelsson, I., Haug, P., & Persson, B. (2005). Inclusive education in some Western European countries: Different policy, rhetorics and school realities. In D. Mitchell (Ed.), *Contextualizing inclusive education. Evaluating old and new international perspectives* (pp. 114–138). London, New York: Routledge.
Emerson, E. (2003). Prevalence of psychiatric disorder in children and adolescents with and without intellectual disabilities. *Journal of Intellectual Disability Research, 47*, 51–58.
* Erickson, K. J. (2000). *The impact of inclusion: A comparison of reading and math achievement of general education middle school students in inclusive versus non-inclusive settings*. Unpublished doctoral dissertation. MI: Graduate State of Wayne State University, Detroit.
Erten, O., & Savage, R. S. (2012). Moving forward in inclusive education research. *International Journal of Inclusive Education, 16*, 221–233.
European Commission. (2014). The structure of the European education systems 2014/2015: Schematic diagrams. Retrieved from http://eacea.ec.europa.eu/education/eurydice/facts_and_figures_en.php#diagrams.
Farrell, P. (2000). The impact of research on developments in inclusive education. *International Journal of Inclusive Education, 4*, 153–162.
Farrell, M., Dyson, A., Polat, F., Hutcheson, G., & Gallannaugh, F. (2007). Inclusion and achievement in mainstream schools. *European Journal of Special Needs Education, 22*, 131–145.
* Feyerer, E. (1998). *Behindern Behinderte? Integrative Unterricht an der Sekundarstufe I*. Innsbruck-Wien: Studien Verlag.
Finn, Ch. E., Manno, B. V., & Vanourek, G. (2000). *Charter schools in action: Renewing public education*. Princeton: Princeton University Press.
* Fletcher, J. (2010). Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management, 29*, 69–83.
Forlin, Ch (2001). Inclusion: Identifying potential stressors for regular class teachers. *Educational Research, 43*, 235–245.
Forlin, Ch (2010). Teacher education reform for enhancing teachers' preparedness for inclusion. *International Journal of Inclusive Education, 14*, 649–653.
Freeman, S. F. N., & Alkin, M. C. (2000). Academic and social attainments of children with mental retardation in general education and special education settings. *Remedial and Special Education, 21*, 3–18.
* Friesen, J., Hickey, R., & Krauth, B. (2009). *Disabled peers and academic achievement*. Research Report. Switzerland: Simon Fraser University.
Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What is intensive instruction and why is it important? *Teaching Exceptional Children, 46*, 13–18.
* Gandhi, A. G. (2007). Context matters: Exploring relations between inclusion and reading achievement of students without disabilities. *International Journal of Disability, Development and Education, 54*, 91–112.
Giangreco, M. F., & Broer, S. M. (2005). Questionable utilization of paraprofessionals in inclusive schools. Are we addressing symptoms or causes? *Focus on Autism and other Developmental Disabilities, 20*, 10–26.
Giangreco, M. F., Broer, S. M., & Suter, J. C. (2011). Guidelines for selecting alternatives to overreliance on paraprofessionals: Field-testing in inclusion-oriented schools. *Remedial end Special Education, 32*, 22–38.
Giangreco, M. F., & Suter, J. C. (2015). Precarious or purposeful? Proactively building inclusive special education service delivery on solid ground. *Inclusion, 3*, 112–131.
Giangreco, M. F., Suter, J. C., & Doyle, M. B. (2010). Paraprofessionals in inclusive schools: A review of recent research. *Journal of Educational and Psychological Consultation, 20*, 41–57.
Göransson, K., & Nilholm, C. (2014). Conceptual diversities and empirical shortcomings — a critical analysis of research on inclusive education. *European Journal of Special Needs Education, 29*, 265–280.
Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis, 36*, 20–43.
* Grimaldi, M. A. (2009). *The academic achievement and engagement of non-disabled male and female students educated within middle-level inclusive and general education classrooms*. Unpublished doctoral dissertation. Oakdale, NY: Dowling College.
* Haeberlin, U., Bless, G., Moser, U., & Klaghofer, R. (1999). *Die Integration von Lernbehinderten*. Bern – Stuttgat: Verlag Paul Haupt.
Hamilton-Jones, B., & Moore, A. (2013). Ensuring high quality inclusive practices: What co- teachers can do. *Kappa Delta Pi Record, 49*, 156–161.
* Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities?. *The Review of Economics and Statistics, 84*, 584–599.
* Hartfield, L. R. (2009). *Mathematics achievements of regular education students by placement in inclusion and non-inclusion classrooms and their principals' perceptions of inclusion*. Unpublished doctoral dissertation. Hattiesburg, MS: The University of Southern Mississippi.
Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
Hattie, J., & Yates, G. (2014). *Visible learning and the science of how we learn*. London: Routledge.
Henderson, K., Klein, S., Gonzalez, P., & Bradley, R. (2005). Teachers of children with emotional disturbance: A national look at preparation, teaching conditions, and practices. *Behavioral Disorders, 31*, 6–17.
* Herrmann, M. A. (2011). *Peer education and peer effects in school*. Research Report. New York, NY: Columbia University.
*. In Hinz, A., Katzenbach, D., Rauer, W., Schuck, K. D., Wocken, H., & Wudtke, H. (Eds.), *Die Entwicklung der Kinder in der Integrativen Grundschule*. Hamburg: Hamburger Buchwerkstatt.
Hollowood, T. M., Salisbury, C. L., Rainforth, B., & Palombaro, M. M. (1994). Use of instructional time in classrooms serving students with and without severe disabilities. *Exceptional Children, 61*, 242–253.
Hoppey, D., & McLeskey, J. (2013). A case study of principal leadership in an effective inclusive school. *Journal of Special Education, 46*, 245–256.
Hornby, G. (2011). Inclusive education for children with special educational needs: A critique. *International Journal of Disability, Development and Education, 58*, 321–329.
Hornby, G. (2014). *Inclusive special education. Evidence-based practices for children with special needs and disabilities*. New York, Heidelberg, Dordrecht, London: Springer.

Hornby, G., & Evans, B. (2014). Including students with significant social, emotional and behavioral difficulties in mainstream school setting. In P. Garner, J. M. Kauffman, & J. Elliott (Eds.), *The SAGE handbook of emotional and behavioral difficulties* (pp. 335–347). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

* Huber, K. D., Rosenfeld, J. G., & Fiorello, C. A. (2001). The differential impact of inclusion and inclusive practices on high, average and low achieving general education students. *Psychology in the Schools, 38*, 497–504.

Jordan, A., Schwartz, E., & McGhie-Richmond, D. (2009). Preparing teachers for inclusive classrooms. *Teaching and Teacher Education, 25*, 535–542.

Kalambouka, A., Farrell, P., Dyson, A., & Kaplan, I. (2007). The impact of placing pupils with special educational needs in mainstream schools on the achievement of their peers. *Educational Research, 49*, 365–382.

Katz, J., Mirenda, P., & Auerbach, S. (2002). Instructional strategies and educational outcomes for students with developmental disabilities in inclusive "multiple intelligences" and typical inclusive classrooms. *Research and Practice for Person With Severe Disabilities, 27*, 227–238.

Kauffman, J. M., Bruce, A., & Lloyd, J. W. (2012). Response to intervention (RtI) and students with EBD. In J. P. Bakken, F. E. Obiakor, & A. Rotatori (Eds.), *Advances in special education, Vol. 23 — behavioral disorders: Current perspectives and issues* (pp. 107–127). Bingley, UK: Emerald.

Kauffman, J. M., & Hallahan, D. P. (Eds.). (2005). *The illusion of full inclusion: A comprehensive critique of a current special education bandwagon* (2nd ed.). Austin, TX: PRO-ED.

Kauffman, J. M., Landrum, T. J., Mock, D. R., Sayeski, B., & Sayeski, K. L. (2005). Diverse knowledge and skills require a variety of instructional groups: A position statement. *Remedial and Special Education, 26*, 2–6.

* Kettmann Klinger, J., Vaughn, S., Hughes, M. T., Schumm, J. S., & Elbaum, B. (1998). Outcomes for students with and without learning disabilities in inclusive classrooms. *Learning Disabilities Research and Practice, 13*, 153–161.

King-Sears, M. E., & Bowman-Kruhm, M. (2011). Specialized reading instruction for adolescents with learning disabilities: What special education co-teachers say. *Learning Disabilities Research & Practice, 26*, 172–184.

Kozleski, E. B., Artiles, A., & Waitoller, F. (2014). Equity in inclusive education: A cultural historical comparative perspective. In L. Florain (Ed.), *The SAGE handbook of special education* (pp. 335–347). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

Kozleski, E. B., & Waitoller, F. R. (2010). Teachers learning for inclusive education: Understanding teaching as cultural and political practice. *International Journal of Inclusive Education, 14*, 655–666.

Kurth, J. A., Morningstar, M. E., & Kozleski, E. B. (2014). The persistence of highly restrictive special education placements for students with low-incidence disabilities. *Research and Practice for Persons with Severe Disabilities, 39*, 227–239.

* LaFever, K. M. (2012). *The effect of co-teaching on student achievement in ninth grade physical science classroom*. Unpublished doctoral dissertation. St. Louis, MO: University of Missouri — St. Louis.

Lindsay, G. (2007). Educational psychology and the effectiveness of inclusive education/mainstreaming. *British Journal of Educational Psychology, 77*, 1–24.

Lipsky, D., & Gartner, A. (1996). Inclusion, school restructuring, and the remaking of American society. *Harvard Educational Review, 66*, 762–797.

Loreman, T. (2007). Seven pillars of support for inclusive education: Moving from "why?" to "how?". *International Journal of Whole Schooling, 3*, 22–38.

Loreman, T. (2014). Measuring inclusive education outcomes in Alberta, Canada. *International Journal of Inclusive Education, 18*, 459–483.

Malinen, O.-P., Savolainen, H., & Xu, J. (2012). Beijing in-service teachers' self-efficacy and attitudes towards inclusive education. *Teaching and Teacher Education, 28*, 526–534.

* Marcheski, J. R. (1997). *The relationship between cooperative teaching in an inclusive setting and the social studies achievement of middle school students without disabilities*. Unpublished doctoral dissertation. Wilmington, OH: Wilmington College.

Matzen, K., Ryndak, D., & Nakao, T. (2010). Middle school teams increasing access to general education for students with significant disabilities: Issues encountered and observations across contexts. *Remedial and Special Education, 31*, 287–304.

* Maultsby-Springer, B. M. (2009). *A descriptive analysis of the impact of co-teaching on the reading/language arts and math achievement of selected middle school students*. Unpublished doctoral dissertation. Nashville, TN: Tennessee State University.

* McCartney, S. P. (2006). *The effects of inclusive school environments on the academic achievement of elementary general education students as measured by standardized tests over a single and multi-year period*. Unpublished doctoral thesis. South Orange, NJ: Seton Hall University.

McCray, E. D., Butler, T. W., & Bettini, E. (2014). Sharing expertise and shifting expectations: The roles of general and special educators in effective, inclusive schools. In J. McLeskey, N. L. Waldron, F. Spooner, & B. Algozzine (Eds.), *Handbook of research for effective, inclusive schools* (pp. 80–93). New York: Routledge, Taylor, & Francis.

McDonnell, J., Johnson, J. W., Polychronis, S., Riesen, T., Jameson, M., & Kercher, K. (2006). Comparison of one-to-one embedded instruction in general education classes with small group instruction in special education classes. *Education and Training in Developmental Disabilities, 41*, 125–138.

* McDonnell, J., Thorson, N., Disher, S., Mathot-Buckner, C., Mendel, J., & Ray, L. (2003). The achievement of students with developmental disabilities and their peers without disabilities in inclusive settings: An exploratory study. *Education and Treatment of Children, 26*, 224–236.

McDonnell, J., Thorson, N., & McQuivey, C. (2000). Comparison of the instructional contexts of students with severe disabilities and their peers in general education classes. *Journal of the Association for Persons With Severe Handicaps, 25*, 54–58.

McLeskey, J., Landers, E., Williamson, P., & Hoppey, D. (2012). Are we moving toward educating students with disabilities in less restrictive settings? *Journal of Special Education, 46*, 131–140.

McLeskey, J., & Waldron, N. (2002). School change and inclusive schools: Lesson learned from practice. *Phi Delta Kappan, 84*, 65–72.

McLeskey, J., & Waldron, N. J. (2006). Comprehensive school reform and inclusive schools. *Theory into Practice, 45*, 269–278.

McLeskey, J., Waldron, N. L., & Reed, L. (2014). A case study of a highly effective, inclusive elementary school. *Journal of Special Education, 48*, 59–70.

McLeskey, J., Waldron, N. L., Spooner, F., & Algozzine, B. (2014). What are effective inclusive schools and why are they important? In J. McLeskey, N. L. Waldron, F. Spooner, & B. Algozzine (Eds.), *Handbook of effective inclusive schools. Research and practice* (pp. 3–16). New York and London: Routledge.

Miles, S., & Singal, N. (2010). The Education for All and inclusive education debate: Conflict, contradiction or opportunity? *International Journal of Inclusive Education, 14*, 1–15.

Mitchell, D. (2014). *What really works in special and inclusive education. Using evidence-based teaching strategies*. London – New York: Routledge.

Mu, G. M. (2015). Pedagogical practices and support systems of inclusion: Empirical evidence from Chinese societies. *International Journal of Disability, Development and Education, 62*, 551–555.

Murawski, W. (2006). Student outcomes in co-taught English classes: How ca we improve? *Reading and Writing Quarterly, 22*, 227–247.

Murawski, W. W., & Lochner, W. W. (2011). Observing co-teaching: What to ask for, look for, and listen for. *Intervention in School and Clinic, 46*, 174–183.

Murawski, W. W., & Swanson, H. L. (2001). A meta-analysis of co-teaching research. Where are the data? *Remedial and Special Education, 22*, 258–267.

* Muscelli, G. B. (2011). *The effect of co-teaching on ninth grade world history students' academic achievement and classroom behavior*. Unpublished doctoral dissertation. Fort Lauderdale, FL: Nova Southeastern University.

* Neugebauer, N. G. (2008). *TAKS scores of general education students in secondary co-teach classes in a Texas school district*. Unpublished doctoral dissertation. College Station, TX: Texas A&M University.

Norwich, B. (2008). *Dilemmas of difference, inclusion, and disability*. Abingdon, Oxon, UK: Routledge.

Norwich, B. (2008b). Dilemmas of difference, inclusion, and disability: International perspectives on placement. *European Journal of Special Needs Education, 23*, 287–304.

Norwich, B. (2014). Categories of special educational needs. In L. Florian (Ed.), *The SAGE handbook of special education* (Vol. 1, pp. 55–71). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

Oh-Young, C., & Filler, J. (2015). A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities. *Research in Developmental Disabilities, 47*, 80–92.

Opertti, R., Walker, Z., & Zhang, Y. (2014). Inclusive education: From targeting groups and schools to achieving quality education as the core of EFA. In L. Florian (Ed.), *Vol. 1. The SAGE handbook of special education* (pp. 149–169). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

* Parker, A. K. (2010). *The impact of co-teaching on general education students*. Unpublished doctoral dissertation. Orlando, FL: University of Central Florida.
Pearce, M., & Forlin, Ch (2005). Challenges and potential solutions for enabling inclusion in secondary schools. *Australasian Journal of Special Education, 29*, 93–105.
Peck, Ch. A., Staub, D., Gallucci, Ch., & Schwartz, I. (2004). Parent perception of the impacts of inclusion on their nondisabled child. *Research and Practice for Persons with Severe Disabilities, 29*, 135–143.
Peebles, J. L., & Mendaglio, S. (2014). The impact of direct experience on preservice teachers' self-efficacy for teaching in inclusive classrooms. *International Journal of Inclusive Education, 18*, 1321–1336.
Peltopuro, M., Ahonen, T., Kaartinen, J., Seppälä, H., & Närhi, V. (2014). Borderline intellectual functioning: A systematic literature review. *Intellectual and Developmental Disabilities, 52*, 419–443.
Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*, 4544–4562.
Pfahl, L., & Powell, J. J. W. (2011). Legitimating the school segregation. The special education profession and the discourse of learning disability in Germany. *Disability & Society, 26*, 449–462.
Powell, J. J. W. (2009). To segregate or to separate? Special education expansion and divergence in the United States and Germany. *Comparative Education Review, 53*, 169–187.
R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press of Harvard University Press.
Rea, P. J., McLaughlin, V. L., & Walther-Thomas, Ch (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children, 68*, 203–222.
Rix, J., Sheehy, K., Fletcher-Campbell, F., Crisp, M., & Harper, A. (2013). Exploring provision for children identified with special educational needs: An international review of policy and practice. *European Journal of Special Needs Education, 28*, 375–391.
Rouse, M., & Florian, L. (2006). Inclusion and achievement: Student achievement in secondary schools with higher and lower proportions of pupils designated as having special educational needs. *International Journal of Inclusive Education, 10*, 481–493.
Ruijs, N. M., & Peetsma, T. T. D. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review, 4*, 67–79.
* Ruijs, N. M., Van der Veen, J., & Peetsma, T. T. D. (2010). Inclusive education and students without special educational needs. *Educational Research, 52*(4), 351–390.
Ryndak, D. L., Taub, D., Jorgensen, Ch. M., Gonsier-Gerdin, J., Arndt, K., Sauer, J., et al. (2014). Policy and the impact on placement, involvement, and progress in general education: Critical issues that require rectification. *Research and Practice for Persons with Severe Disabilities, 39*, 65–74.
* Saint-Laurent, L., Dionne, J., Glasson, J., Royer, E., Simard, C., & Pierard, B. (1998). Academic achievement effects of an in-class service model on students with and without disabilities. *Exceptional Children, 64*, 239–253.
Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children, 73*, 392–416.
* Sermier Dessemontet, R., Benoit, V., & Bless, G. (2011). Schulische integration von kindern mit einer geistigen behinderung - untersuchung der entwicklung der schulleistungen und der adaptiven fähigkeiten, der wirkung auf die lernentwicklung der mitschüler sowie der lehrereinstellungen zur integration. *Empirische Sonderpädagogik, 4*, 291–307.
Sermier Dessemontet, R., & Bless, G. (2013). The impact of including children with disability in general education classrooms on the academic achievement of their low-, average-, and high-achieving peers. *Journal of Intellssectual & Developmental Disability, 38*, 23–30.
* Sharpe, M. N., York, J. L., & Knight, J. (1994). Effects of inclusion on the academic performance of classmates without disabilities. *Remedial and Special Education, 15*(5), 281–287.
* Shinn, M. R., Powell-Smith, K. A., Good, R. H., III, & Baker, S. (1997). The effects of reintegration into general education reading instruction for students with mild disabilities. *Exceptional Children, 64*(1), 59–79.
Slee, R. (2001). Social justice and the changing directions in educational research: The case of inclusive education. *International Journal of Inclusive Education, 5*, 167–177.
Solis, M., Vaughn, S., Swanson, E., & McCulley, L. (2012). Collaborative models of instructions: The empirical foundations of inclusion and co-teaching. *Psychology in the Schools, 49*, 498–510.
Soukakou, E. P. (2012). Measuring quality in inclusive preschool classrooms: Development and validation of the Inclusive Classroom Profile (ICP). *Early Childhood Research Quarterly, 27*, 478–488.
Stanovich, P. J., & Jordan, A. (1998). Canadian teachers' and principals' beliefs about inclusive education as predictors of effective teaching in heterogeneous classrooms. *The Elementary School Journal, 98*, 221–238.
* Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal, 32*, 321–351.
Szumski, G., & Karwowski, M. (2014). Psychosocial functioning and school achievement of children with mild intellectual disability in Polish special, integrative, and mainstream schools. *Journal of Policy and Practice in Intellectual Disabilities, 11*, 99–108.
Taddlie, C., Kirby, P. C., & Stringfield, S. (1989). Effective versus ineffective schools: Observable differences in the classroom. *American Journal of Education, 97*, 221–236.
Talmor, R., Reiter, S., & Feigin, N. (2005). Factors relating to regular education teacher burnout in inclusive education. *European Journal of Special Needs Education, 20*, 215–229.
Theoharis, G., & Causton-Theoharis, J. (2011). Preparing pre-service teachers for inclusive classrooms: Revising lesson-planning expectations. *International Journal of Inclusive Education, 15*, 743–761.
* Thompson, K. (2010). *The many faces of co-teaching: How does co-teaching impact students at different levels of academic functioning?*. Unpublished doctoral dissertation. Carrollton, GA: University of West Georgia.
* Trabucco, M. (2011). *The influence of co-taught inclusion on the academic achievement of third grade non-disabled students in mathematics*. Unpublished doctoral dissertation. South Orange, NJ: Seton Hall University.
* Trejo, Ch. L. (2008). *The impact of inclusion programs on the academic achievement of non-disabled learners in selected Texas elementary schools*. Unpublished doctoral dissertation. Stephenville, TX: Tarleton State University.
UN (United Nations). (2006). *Convention on the rights of persons with disabilities*. New York, NY: United Nations.
UNESCO. (1994). *The Salamanca statement and framework for action on special needs education*. Paris: Author.
UNESCO. (2012). *International standard classification of education*. Montreal, Canada: United Nations.
Valentine, J. C. (2009). Judging the quality of primary research for research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 129–146). New York: Russell Sage Foundation.
Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children, 67*, 99–114.
van der Veen, I., Smeets, E., & Derriks, M. (2010). Children with special educational needs in The Netherlands: Number, characteristics, and school career. *Educational Research, 52*, 15–43. http://dx.doi.org/10.1080/00131881003588147.
Verdonschot, M. M. L., de Witte, L. P., Reichrath, E., Buntinx, W. H. E., & Curfs, L. M. G. (2009). Community participation of people with intellectual disability: A review of empirical findings. *Journal of Intellectual Disability Research, 53*, 303–318.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

de Vroey, A., Struyf, E., & Petry, K. (2016). Secondary schools included: A literature review. *International Journal of Inclusive Education, 20*, 109–135.
Wagner, M., Friend, M., Bursuck, W. D., Kutash, K., Duchnowski, A. J., Sumi, W. C., et al. (2006). Educating students with disabilities: A national perspective on programs and services. *Journal of Emotional and Behavioral Disabilities, 14*, 12–30.
Wang, M. C., & Baker, E. T. (1985–1986). Mainstreaming programs: Design features and effects. *The Journal of Special Education, 19*, 503–521.
Weiss, M. P., & Lloyd, J. W. (2002). Congruence between roles and actions of secondary special educators in co-taught and special education setting. *The Journal of Special Education, 36*, 58–68.
* Whisted, M. L. (2011). *Does the use of co-teaching models in algebra result in an increase in student achievement among students with disabilities and their non-disabled peers?*. Unpublished doctoral dissertation. Balitimore, MD: College of Notre Dame of Maryland.
* Willrodt, K., & Claybrook, S. (1995). *Effects of inclusion on academic outcomes*. Unpublished doctoral dissertation. Huntsville, TX: Sam Houston State University.
Wilson, D. B. (2014). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html.
* Wocken, H. (1987). Schulleistungen inIntegrationklasse. In H. Wocken, & G. Antor (Eds.), *Integrationsklassen in Hamburg* (pp. 276–308). Solms-Oberbiel: Jarick.
Zigmond, N. (2006). Reading and writing in co-taught secondary school social studies classrooms: A reality check. *Reading and Writing Quarterly, 22*, 249–268.