Intelligence 88 (2021) 101566

Contents lists available at ScienceDirect

Intelligence

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

A valid evaluation of the theory of multiple intelligences is not yet possible: Problems of methodological quality for intervention studies

Marta Ferrero [a,**], Miguel A. Vadillo [b], Samuel P. León [c,*]

[a] Facultad de Educación, Universidad Complutense de Madrid, Madrid, Spain
[b] Departamento de Psicología Básica, Universidad Autónoma de Madrid, Madrid, Spain
[c] Departamento de Pedagogía, Universidad de Jaén, Jaén, Spain

ARTICLE INFO

Keywords:
Intervention
Multiple intelligences
Systematic review
Meta–analysis

ABSTRACT

Since Gardner suggested that human beings hold multiple intelligences, numerous teachers have adapted and incorporated the multiple intelligence theory (MIT) into their daily routine in the classroom. However, to date, the efficacy of MIT–inspired methodologies remains unclear. The focus of the present study was to perform a systematic review and a meta–analysis to assess the impact of these interventions on academic achievement through reading, maths, or science tests. The inclusion criteria for the review required that studies should estimate quantitatively the impact of an MIT–based intervention on the academic performance and that they followed a pre–post design with a control group. The final sample included 39 articles comprising data from 3009 pre-school to high school students, with diverse levels of achievement, from 14 different countries. The results showed that the studies had important methodological flaws, like small sample sizes or lack of active control groups; they also reported insufficient information about key elements, such as the tools employed to measure the outcomes or the specific activities performed during training, and revealed signs of publication or reporting biases that impeded a valid evaluation of the efficacy of MIT applied in the classroom. The educational implications of these results are discussed.

1. Introduction

In 1983 Howard Gardner published *Frames of Mind*, where he laid out, for the first time, the idea that human beings possess not one but multiple intelligences, each one defined as the ability to solve a specific problem or create a product which is perceived as valuable in one or more context–rich settings. According to the initial proposal of the author, each of these cognitive competencies were independent of each other, so that the same individual might be strong in one intelligence but weak in another one. Gardner (1993) established eight basic criteria to identify an intelligence: (a) potential isolation by brain damage; (b) the existence of savants, prodigies, and other exceptional people; (c) a distinctive developmental history; (d) a definable set of expert *end–state* performances or professions; (d) support from psychometric findings; (e) support from experimental psychological tasks; (f) an identifiable set of operations; and (g) susceptibility to encoding in a symbol system. Following these criteria, he initially proposed the existence of seven intelligences: Linguistic, logical mathematical, musical, spatial, bodily kinaesthetic, interpersonal, and intrapersonal. Fourteen years later, the original list was enriched with a new one, naturalist intelligence (Gardner, 1997).

Along with this new approach to the conceptualization of human intelligence, Gardner also put forward a new way to assess it. In his view, the prevailing methods at the time were exclusively focused on the measurement of linguistic and logical capacities. In addition, they usually consisted in paper–and–pencil isolated tasks, detached from any culture and frequently unfamiliar to the children (Gardner, 1993). To overcome these limitations, Gardner suggested what he called intelligence–fair measures. They consisted in a number of culturally meaningful activities, always related to particular professions, which enabled the assessment of the psychological processes inherent to each intelligence in in one– to two–hour sessions. This proposal would contribute to demonstrate the existence of independent intelligences and, hence, to identify the strengths and weaknesses of individuals (Gardner, 1983).

* Correspondence to: S. P. León, Departamento de Pedagogía, Campus de las Lagunillas s/n, 23071 Jaén, Spain.
** Correspondence to: M. Ferrero, Departamento de Investigación y Psicología de la Educación, c/ Rector Royo Villanova s/n Ciudad Universitaria, 28040 Madrid, Spain.
E-mail addresses: maferr16@ucm.es (M. Ferrero), sparra@ujaen.es (S.P. León).

https://doi.org/10.1016/j.intell.2021.101566
Received 26 May 2020; Received in revised form 1 July 2021; Accepted 2 July 2021
Available online 19 July 2021

Due to the high costs of the fair measures in terms of money and effort, Gardner would later propose a new tool to measure intelligence, the Modified Spectrum Field Inventory, which sampled several intelligences in two one–hour sessions (Gardner & Hatch, 1989).

The Multiple Intelligences Theory (MIT) was not originally conceived to be directly applied to educational settings (Gardner, 1983). However, Gardner has always declared himself convinced of its potential contribution to this field. It was his understanding that schools have traditionally assumed everyone can learn the same contents in the same way; therefore, they have not tried to associate learning experience to specific products in the life of the community to enhance a rich understanding of knowledge. Additionally, schools have tended to cultivate almost exclusively the linguistic and logical–mathematical symbolization and have done so through repetitive and rote learning (Gardner, 1991). As a result, many children fail to discover their gifts or talents (Gardner & Hatch, 1989). According to Gardner (2011), the educational system can be notably improved by encouraging teachers to consider the different preferences and intellectual strengths of students and teach them in ways consistent with them. In this vein, the adoption of a MIT curriculum becomes a promising option to foster learning. Although the author has always defended that MIT can be applied in numerous manners, he has suggested some specific means to exercise and develop intelligences (Gardner, 1991). More precisely, the author has highlighted the importance of creating opportunities to work intensively with rich and engaging materials that include societal roles and enhance the involvement of different human intelligences. To achieve this aim, the classroom is organised in different corners or areas where children can freely explore the learning materials and connect with the respective set of intelligences involved in them. For instance, a classroom may include a naturalistic corner, where students might examine and compare biological specimens with other materials (Gardner, 1991). This kind of intervention allows teachers to identify children's talents and unexpected strengths without any specific assessment and to promote in them attitudes and basic skills through the use of resources in which they have demonstrated an interest and emerging expertise. Faced with a plethora of poor pedagogical proposals supposedly inspired by the theory, Gardner has also established a firm distinction between positive and *not right* approaches to applying his theory (Gardner, 1995) and has highlighted the value of various educational projects as an example of good practice (Gardner, 1997). In addition, he has collaborated in the publication of different books on how to apply MIT in the classroom (Chen, Feldman, & Gardner, 1998a, 1998b, 1998c; Chen, Moran, & Gardner, 2009) and shown support to related works of other authors (Amstrong, 2009). Finally, he is currently leading Project Spectrum, an initiative aimed at offering a new approach for assessment and curriculum development in schools (Harvard Graduate School of Education, 2016).

To this day, the empirical evidence supporting the existence of multiple unrelated intelligences is weak. The core elements of the MIT have been criticised on countless occasions (Geake, 2008; Waterhouse, 2006; White, 2004; Willingham, 2004; but see Gardner & Morgan, 2006). In opposition to Gardner's assertion about the existence of eight independent intelligences, the scientific community closes ranks around the hierarchical nature of intelligence and the existence of a general factor which explains a notable percentage of individual differences (Colom, 2018; Hunt, 2001; Jensen, 1998; Lubinski, 2004; Visser, Asthon, & Vernon, 2006a). Furthermore, a recent study has shown that many of the tests aimed at measuring the different domains in Gardner's framework are not only strongly intercorrelated with each other but also with external tests of general intelligence (Visser, Asthon, & Vernon, 2006b; but see Gardner, 2006).

Beyond the plausibility of MIT, the aim of the present work is to assess its impact on schools. Since its formulation, MIT has attracted the enthusiasm and interest of a growing number of teachers around the world (White, 2004). Although Gardner (1983) himself acknowledged that he had employed the term intelligence, and not talent or ability, to grab the attention of the audience, the idea that all people could be smart in some way and that educators should ensure that all intelligences are equally exercised among students has been embraced wholeheartedly by large parts of the educational community and inspired the creation and introduction of a considerable number of methodologies and resources in schools (Amstrong, 2009; Collin, 2001; Gardner, 1997). At the same time, MIT has stimulated the completion of several studies aimed at measuring the impact of the theory on the academic performance of students (Bas, 2016; Batdi, 2017). In general, these studies have analysed the effect of MIT intervention through the use of MIT-labelled activities and materials on outcomes such as science, reading or mathematics, compared to more traditional methods. In most of the cases, all the intelligences have been addressed in each learning session. The results collected to date suggest that, in general, students trained with MIT intervention outperform control groups. However, as detailed below, not all the studies have found a statistically significant benefit for MIT-based interventions. These contradicting results might be partially explained by methodological differences among the studies, such as the size of the groups or the inclusion of a control group. On the latter, just one study included an active control group (Modirkhamene & Azhiri, 2012). The authors of the study checked the effect of MIT-based intervention on reading comprehension of a sample of 70 secondary students during two months. In spite of the good results obtained in favor of the experimental group, the inclusion of various practices labelled as non-valid by Gardner (1995) himself makes it unfeasible to consider this work a good estimate of what to expect in this area. Given the importance of basing educational practices on solid empirical evidence, the focus of the present article is precisely to perform a systematic review and meta-analysis that allow assessing the impact of MIT–inspired instructional methodologies on the academic achievement of learners after a careful examination of any bias or methodological issues.

1.1. Overview of the present systematic review

Until now, two meta–analyses have explored the impact of MIT on the academic performance of students. From our point of view, both of them present important methodological shortcomings that preclude their use as a solid reference to inform educational practice. The first meta–analysis (Bas, 2016) included only master theses ($k = 64$) and doctoral dissertations ($k = 11$) published in Turkey between 1998 and 2014. The second meta–analysis (Batdi, 2017) consisted of 63 articles and doctoral dissertations conducted around the world between 2000 and2016. Both meta–analyses applied minimal quality criteria for the selection of studies and made no attempt to measure or control the risk of bias induced by different aspects of the designs and procedures, such as blinding or the use of passive control conditions. In the same vein, none of them analysed the potential impact of publication bias or selective reporting on the results. Given these shortcomings, it is perhaps unsurprising that both meta–analyses obtained remarkably large average effects ($d = 1.077$ and 0.95, respectively). Based on such limited evidence, it is impossible to assess whether these large effects should be attributed to genuine MIT–based interventions or to the multiple sources of bias that could influence the results of the individual studies included in the meta–analyses. In addition, the procedure and criteria used to search for primary studies were not defined with sufficient detail to reproduce the results and none of the reviews offered a list of included and excluded studies that could be used to confirm and extend their analyses.

In contrast to previous reviews, the literature search strategy adopted in the present study aimed at locating both studies published in peer–reviewed journals through the Web of Science and also grey literature through ProQuest and Google Scholar, by using a well–defined procedure that any reader with access to these databases will be able to reproduce. Secondly, one of our main goals was to assess the quality of each individual study and to identify potential sources of bias. Finally,

Table 1
Articles that met inclusion and quality criteria.

Author, year	Country	Sample type	N (E / C)	Age	Educational level	Instructor	Duration	Dependent variable	Tests to measure DV	Test to measure MI	Results
Abdi, Laei, & Ahmadyan, 2013	Iran	Normal population	40 (20 / 20)	n.s.	5th grade	Teacher	8 weeks	Science achievement	Achievement test (developed by the teacher)	–	MIT group outperformed significantly control group
Abdulkader, Gundogdu, & Mourad, 2009	Egypt	Learning disabled students	60 (30 / 30)	11–12 years	5th grade	Researcher	2 months, 3 weekly sessions (40–45 min per session)	Decoding and written comprehension skills	Word Recognition test and Reading comprehension Test	–	MIT group outperformed significantly control group
Akkuzu & Akçay, 2011	Turkey	Normal population	75 (38 / 37)	n.s.	High school	Researcher	8 weeks	Periodic feature' variation	Periodic Features' Variation Achievement Test (PFVAT)	Multiple Intelligences Assessment Survey (MIAS)	There are improvements in both groups. No comparison made
Alavinia & Farhady, 2012	Iran	Normal population	60 (30 / 30)	15–20 years	Intermediate education	n.s.	17 sessions (90 min per session)	Vocabulary learning in English	The Preliminary English Test	Name no specified	There are improvements in both groups. No comparison made.
Al-Balhan, 2006	Kuwait	Low academic achievement students	410 (210 / 200)	n.s.	1st, 2nd, 3th and 4th secondary education	Teacher	n.s.	Reading skills	Reading test	Adaptation of Gardner's Multiple Intelligences Inventory (McKenzie, 1999)	MIT group outperformed significantly control group
Alqatanani, 2017	Jordania	Normal population	59 (30 / 29)	n.s.	10th grade	Researcher	2 months, 24 sessions	Reading skills in English	Reading test	–	MIT group outperformed significantly control group
Altıntaş & Özdemir, 2015	Istanbul	Normal population and gifted students	117 (57 / 60)	n.s.	5th and 6th grade	n.s.	n.s.	Mathematical concepts	Mathematics Achievement Test	Multiple Intelligences Inventory (Saban, 2005)	MIT group outperformed significantly control group
Al-ZoubiI & Al-Adawi, 2019	Oman	Students with dyscalculia	14 (7 / 7)	8–10 years	3th and 4th grade	Researcher	10 weeks (40 min 5 lessons per week)	Mathematics achievement	Mathematics Test	The Multiple Intelligences Checklist (Nofal, 2010)	MIT group outperformed significantly control group
Anaduaka, 2008	Nigeria	Normal population	118 (59 / 59)	n.s.	Senior secondary students	Teacher	5 weeks	Geometry achievement	Geometry achievement test (developed by the researcher)	–	MIT group outperformed significantly control group
Bilgin, 2006	Turkey	Low socio-economic status students	50 (25 / 25)	14–16 years	9th grade	Teacher	3 weeks (three times per week)	Chemical knowledge	Chemical bonding achievement test (developed by the researcher)	–	MIT group outperformed significantly control group
Delgoshaei & Delavari, 2012	Iran	Normal population	40 (20 / 20)	n.s.	Preschool education	n.s.	n.s.	Sequential thinking, problem solving skills, basic concepts formation, memory and observation skills, the five senses	Cognitive development questionnaire	–	MIT group outperformed significantly control group
Dillihunt & Tyler, 2006	United States	Low income students	213 (n. s. / n. s.)		3th and 5th grade	Teacher	7 weeks	Multiplication skills	Multiplication test	–	MIT group outperformed significantly control group
	Malaysia	Normal population		n.s.		n.s.	2 months	Writing ability			

(continued on next page)

Table 1 (*continued*)

Author, year	Country	Sample type	N (E / C)	Age	Educational level	Instructor	Duration	Dependent variable	Tests to measure DV	Test to measure MI	Results
Eng & Mustapha, 2010			58 (28 / 30)		Secondary education				Six-Trait Analytic Writing Rubric	The Malaysian Adolescent Multiple-Intelligences Test (MAMIT)	MIT group outperformed significantly control group
Fasni, Fatimah, & Yulanda, 2017	Indonesia	n.s.	63 (32 / 31)	n.s.	n.s.	n.s.	n.s.	Maths problem solving ability	Test of mathematical problem solving ability	The Scale of the Multiple Intelligences Attitude (adapted from Santrock, 2007)	MIT group outperformed significantly control group
Gün & Sahin, 2014	Turkey	Normal population	71 (37 / 34)	n.s.	5th grade	Researcher	4 weeks	Knowledge about social contents	Social Studies Achievement Test	–	MIT group outperformed significantly control group
Gündüz & Üna, 2016	Turkey	Normal population	50 (25 / 25)	11–12 years	6th grade	Teacher	3 weeks (three hours per week)	English writing development	Two essays	–	MIT group outperformed significantly control group
Gurbuz, Birgin, & Catlioglu, 2014	Turkey	Normal population	48 (24 / 14)	n.s.	7th grade	Teacher	2 weeks (4 h per week)	Maths knowledge	Conceptual Learning Test	–	MIT group outperformed significantly control group
Gurcay & Ferah, 2017	Turkey	Normal population	95 (45 / 50)	14–16 years	9th grade	Researcher (E) and teacher (C)	6 weeks	Force and motion knowledge	Force and Motion Achievement Test	Revised Student Multiple Intelligences Profile Questionnaire (SMIP-24) (Chan, 2006)	MIT group outperformed significantly control group
Haboush, 2010	Gaza	Normal population	97 (65 /32)	11–14 years	8th grade	Researcher	6 weeks	Reading comprehension	n.s.	–	MIT group outperformed significantly control group in some of the skills tested
Hanley, Hermiz, Lagioia-Peddy, & Levine-Albuck, 2002	EEUU	Normal population	n.s.	n.s.	5th grade	Teacher and researcher	14 weeks (two days per week for one hour each day)	Achievement in social studies	Matching questions, multiple-choice questions, and an essay question	The MI Test	MIT group outperformed significantly control group
İnan & Erkus, 2017	Turkey	Normal population	64 (32 /32)	n.s.	4th grade	n.s.	n.s.	Mathematics achievement	Achievement test (developed by the researcher)	–	MIT group outperformed significantly control group
Işık & Tarım, 2009	Turkey	Normal population	150 (E1 37, E2 34 / C1 40, C2 39)	n.s.	4th grade	Researcher (E) and teacher (C)	E 21 weeks / C 12 weeks	Maths knowledge	Mathematics Achievement Test (MAT)	Teele Inventory for Multiple Intelligences (TIMI) (Teele, 2000)	MIT group outperformed significantly control group
Kaya, Dogan, Gokcek, Kilic, & Kilic, 2007	Turkey	Normal population	60 (30/ 30)	13–14 years	8th grade	Researcher	4 weeks (three days per week)	Science achievement	Achievement test (developed by the researcher)	Adaptation of Multiple Intelligences Survey of Armstrong, 1994	MIT group outperformed significantly control group
Khalghollah, Afsha, & Shahidi, 2014	Iran	Normal population	66 (33 / 33)	7–8 years	7th and 8th grade	n.s.	8 weeks	Science knowledge	Learning Sciences questionnaires / Science Learning lessons	–	MIT group outperformed significantly control group
	Indonesia	Normal population		n.s.	11th grade	n.s.	n.s.	Science knowledge			

(*continued on next page*)

Table 1 (*continued*)

Author, year	Country	Sample type	N (E / C)	Age	Educational level	Instructor	Duration	Dependent variable	Tests to measure DV	Test to measure MI	Results
Liliawati, Purwanto, Zulfikar, & Kamal, 2018			30 (30 / 30)						The Concept Mastery Test	Multiple Intelligences Questionnaires (Xie & Lin, 2009)	There are improvements in both groups. No comparison made
Modirkhamene & Azhiri, 2012	Iran	Normal population	70 (35 / 35)	16–23 years	EFL intermediate education	n.s.	2 months, 24 sessions (3 sessions of 90 min per week)	Reading comprehension skills	Reading proficiency test	Multiple Intelligences-based Profiling (Armstrong, 1994)	MIT group outperformed significantly control group
Mostafa, 2017	Egypt	Neurologically deficient population and poor reading skills	40 (20 / 20)	n.s.	1th grade	Researcher	10 weeks (3 sessions of 45 min per week)	Reading skills	Moursd Ali's Basic Reading Skills Test	–	MIT group outperformed significantly control group
Nakhbi & Barza, 2016	Arab Emirates	Normal population	53 (27 / 26)	10–12 years	6th grade	Teacher	4 weeks (30 min per session)	Science knowledge	Science content exam	MI survey (Candler, 2011)	Achievement of experimental and control group is not analysed
Nuallaong, Nuallaong, & Preechadirek, 2015	Thailand	Normal population	62 (31 / 31)	n.s.	1th grade	n.s.	n.s.	Vocabulary	Vocabulary test	Teacher assisted questionnaire about MI	MIT group outperformed significantly control group
Özdermir, Güneysu, & Tekkaya, 2006	Turkey	Normal population	70 (35 / 35)	9–10 years	4th grade	Teacher	7 weeks, 6 lessons (45 min per session)	Science knowledge	Diversity of Living Things Concepts Test	Teele Inventory of Multiple Intelligences Test (TIMI)	MIT group outperformed significantly control group
Pahlavani, Khosravani, & Zanjani, 2017	Iran	Normal population	43 (27 / 16)	18–37 years	Undergraduate education	Teacher	8 session (1 h and 15 min per session)	Speaking ability	Nelson Placement Test	Gardner's Multiple Intelligence Questionnaire	MIT group outperformed significantly control group
Safranj & Zivlak, 2018	Serbia	Normal population	58 (30 / 28)	n.s.	Undergraduate education	n.s.	One semester	Language skills	Students' language knowledge	–	MIT group outperformed significantly control group
Sánchez-Martín, Álvarez-Gragera, Dávila-Acedo, & Mellado, 2017	Spain	Normal population	160 (87 / 73)	12–13 years	n.s.	n.s.	n.s.	Technology education	The exam of the topic	Survey version designed by Giorgis, 2007	MIT group outperformed significantly control group
Sheahan, While, & Bloomfield, 2015	Ireland	Normal population	90 (46 / 44)	18 to 51 years	Nursing students	Researcher	12 weeks	Clinical skills	Objective structured clinical examination	A multiple intelligences development assessment scale (MIDAS)	MIT group outperformed significantly control group
Soleimani, Moinnzadeh, Kassaian, & Z., & Ketabi, S., 2012	Iran	Normal population	61 (32 / 29)	n.s.	Undergraduate students	Researcher	8 weeks	English skills	Achievement test (developed by the teacher)	–	MIT group outperformed significantly control group
Stanciu, Orban, & Bocos, 2011	Romania	Learning disabled students	36 (18 / 18)	n.s.	3th and 4th grade	n.s.	8 weeks	Science knowledge	Assessment test	Checklist (http://www.spannj.org/BasicRights/appendix_b.htm)	MIT group outperformed significantly control group
Ucak, Bag, & Usak, 2006	Turkey	Normal population	54 (27 / 27)	12–14 years	7th grade	Teacher	4 weeks	Chemistry knowledge	Chemistry Achievement Test (CACT)	–	There are improvements in both groups. No comparison made

(*continued on next page*)

Table 1 (*continued*)

Author, year	Country	Sample type	N (E / C)	Age	Educational level	Instructor	Duration	Dependent variable	Tests to measure DV	Test to measure MI	Results
Yalmanci & Gozum, 2013	Turkey	Normal population	60 (30 / 30)	n.s.	Undergraduate education	Researcher	4 weeks	Science knowledge	Success test	–	There are improvements in both groups. No comparison made
Young, 2003	Australia	Normal population	44 (22 / 22)	12–13 years	8th grade	Researcher	9 weeks (for four lessons per week per class)	Mathematics achievement	Mathematics test	¿?	There are not significant differences between MIT group and control group

Note: DV: Dependent variable. MI: Multiple Intelligences. MIT: Multiple Intelligences Theory. (E) Experimental Group. (C) Control Group. (E1) Experimental Group I. (E2) Experimental Group II. (C1) Control Group I. (C2) Control Group II. n.s.: Not specified.

unlike previous meta–analyses in this domain, we attempted to detect and measure the impact of publication and reporting biases.

2. Method

2.1. Search procedures

This systematic review follows the recommendations of PRISMA (Moher, Liberati, Tetzlaff, & Altman, 2009) and APA's (Appelbaum et al., 2018) reporting standards for systematic reviews and meta-analyses. On May 1st 2019, the first author (MF) conducted a search on the Web of Science with the term *multiple intelligences* and on August 20th 2020, she repeated the search on ProQuest and Google Scholar with the free software *Publish or Perish* (Harzing, 2007). These searches were limited to English–language articles published after 1983 (the year in which Gardner published *Frames of Minds*). After removing 18 duplicates, the first search returned 937 studies on the Web of Science while the second search returned 1642 studies on ProQuest and 944 studies on Google Scholar.

Titles and abstracts of the studies were screened for eligibility by MF. Studies were only included if they fulfilled the following inclusion criteria: (c1) the aim was to measure quantitatively the impact of an MIT–based intervention on the academic performance of students, and (c2) the study followed a pre–post design with control group. This step resulted in a total of 38 full–text articles on the first search (Web of Science) and 232 full-text articles on the second search (197 from ProQuest and 35 from Google Scholar). Then, MF and SPL independently read these articles to verify that they met the inclusion criteria. As a result of this screening, a set of 15 articles were selected on the first search. Thereupon, descendancy searches of articles citing or cited by these 15 papers were conducted and 129 new studies were identified. As in the previous stage, the titles and abstracts of these articles were screened by MF, resulting in 26 additional full–text publications that were independently read by MF and SPL. Eleven of them complied with the inclusion criteria. Therefore, the sample of articles reviewed for inclusion on the first search consisted of 26 studies (15 + 11). Likewise, the screening of the 232 full–text publications obtained on the second search resulted in the selection of 13 articles (1 from ProQuest and 12 from Google Scholar). Therefore, the final sample of articles reviewed for inclusion consisted of 39 studies (26 + 13). Table 1 presents a summary of the most relevant information of these articles. A PRISMA flowchart summarizing the literature search process conducted in this study is shown in Fig. 1. Across all full–text articles assessed for inclusion, MF and SPL reached an initial inter–rater agreement of 94.64%. Disagreements were discussed until 100% consensus was reached.

2.2. Data extraction and coding

MF and SPL independently coded each of the 39 selected studies, including the title of the journal, size and characteristics of the sample, type of experimenter (researcher or teacher), duration of the intervention, dependent variables, and tests employed to measure them (see Table 1).

As mentioned above, one of the main contributions of this work lies in the assessment of the quality of the studies and the identification of potential sources of bias. To achieve this, we elaborated a 17–item quality scale. For a detailed description of each item, see Table 2. Items 2–3 (randomisation), 4–6 (blinding), 7–8 (similarities between groups), 11 (type of control group, active or passive), 12–13 (information for replicability), 14 (reliability of tests), 15 (validity of tests), and 16 (availability of statistic data) were partially based on the scale developed by Physiotherapy Evidence Database (PEDro) and an educational scale made by Newman (2003) on the basis of CRD guidelines, EPOC Handbook and the Campbell Collaboration Research Design Policy Brief. Items 1 (study preregistration), 9 (training of instructor), 10 (intervention fidelity) and 17 (open access publication) were elaborated ad hoc

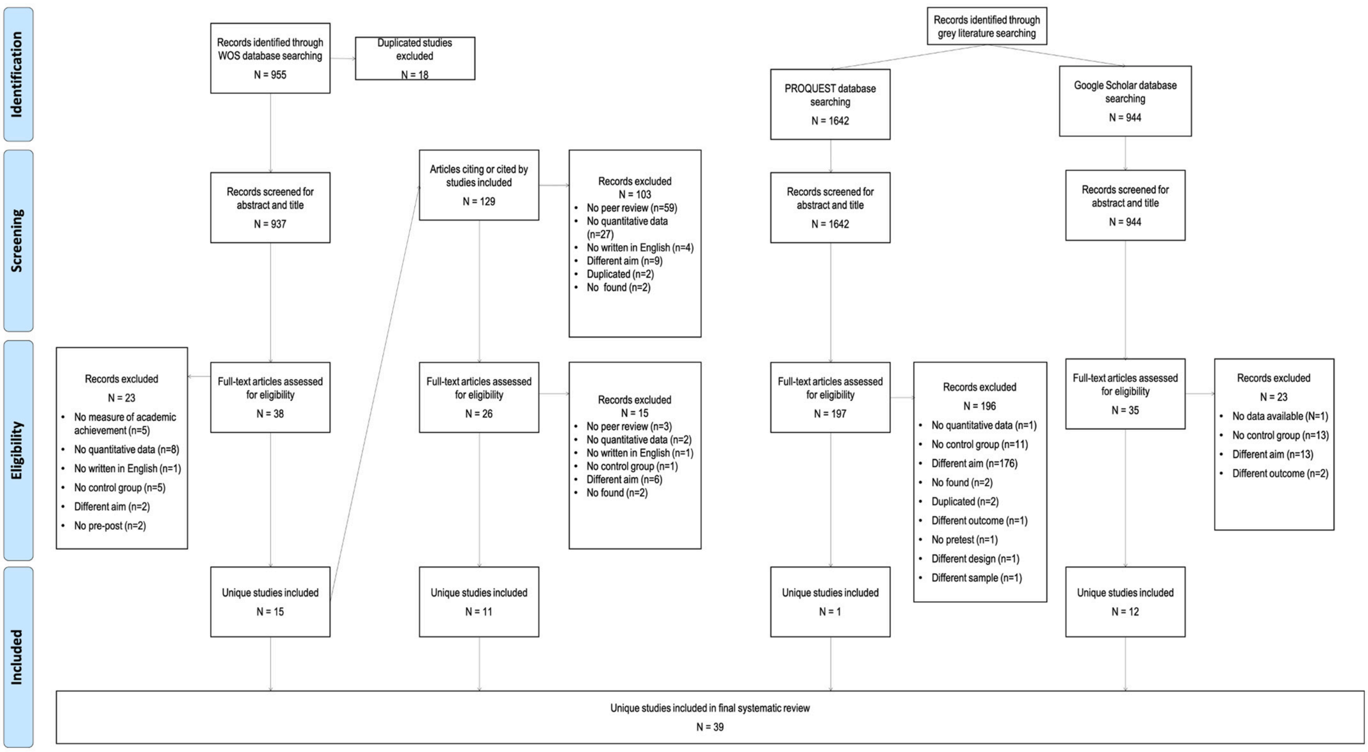

Fig. 1. PRISMA flowchart.

for this study. Each item could be assigned three values: (a) positive, when the study met the criterion; (b) negative, when the study did not meet the criterion; and (c) unknown, when the study did not provide sufficient information. The items that comprise the quality assessment along with the final values assigned to each of them within each study are shown in Fig. 2. For each study, MF and SPL independently assigned a value to each item, reaching an initial agreement of 88.5%. Where disagreements took place, these were resolved through discussion until 100% consensus was reached. The final values reached by each item after assessing all the studies are described in Fig. 3. Following criticisms to the use of quality scores (Jüni, Witschi, Bloch, & Egger, 1999), we did not attempt to aggregate the items of the scale into a single score.

2.3. Computation of effect sizes and statistical analyses

In addition to collecting the qualitative data reported in the previous paragraph, we also extracted information to estimate the effect size of the intervention in each study. Given that all the studies included in the systematic review were controlled trials with pre– and post–test scores, we used the standardized mean change difference score recommended by Morris (2008) to quantify the magnitude of the effect in studies with this type of design. However, some studies did not report sufficient information about participants' performance in the pre–test to compute this type of effect size. Therefore, we also conducted a secondary meta–analysis using the standardized mean difference of post–test scores.

Standardized mean change difference scores, here denoted as g_Δ, were computed with Eqs. 8–10 from Morris (2008) and their variance was computed following Eq. 25 from the same source. In these equations, only the standard deviation of pre–test scores in both groups is used to standardize the mean change difference. The computation of the variance of g_Δ requires an estimate of the correlation between pre– and post–test scores, a piece of information that is usually missing in most studies. Following Rosenthal (1991), we assumed a correlation of 0.70 between pre– and post–test scores in all the analyses reported in the present article. However, we also conducted sensitivity analyses assuming lower (0.50) and higher (0.90) values that gave rise to similar results which, for the sake of simplicity, are not reported.

As explained above, some studies did not report sufficient information to compute g_Δ, but they did report sufficient information to compute a standardized mean difference of post–test scores. Therefore, we conducted an additional meta–analysis with this alternative effect size estimate, which we will denote as g_p. This effect size was computed using Eqs. 4.18 and 4.19 from Borenstein, Hedges, Higgins, and Rothstein (2009) including also the correction factor J computed with Eqs. 4.22 and 4.23. The variance of g_p was computed using Eqs. 4.20 and 4.24 from the same source.

Some of the studies included in the systematic review measured more than one valid outcome. In these cases, we aggregated all the effect sizes from a single study using the method suggested by Borenstein et al. (2009), as implemented in the agg function of the 'MAd' package for R, assuming a (default) correlation between effect sizes of 0.50. Unless noted otherwise, all the statistical analyses were performed with the 'metafor' package for R (Viechtbauer, 2010) using random-effects models.

3. Results

3.1. Description of the studies

Table 1 provides a summary of the 39 studies included in the present review. Overall, the studies were highly heterogeneous in terms of sample type (e.g. academic achievement or socio–economic level), sample size (from 14 to 410 participants), educational level (from preschool to undergraduate education), duration of the interventions (from two to ten weeks), and type of outcomes (e.g. achievement in reading, mathematics, or science).

3.2. Qualitative assessment

Fig. 2 presents a detailed summary of the qualitative assessment of the 39 studies included in the systematic review. As can be seen, most articles fail to report sufficient information to assess the quality of the

Table 2
Quality scale: Description of the items.

Quality criteria	Items description
Study pre-registration	1. The rationale, hypotheses, design, and analytic strategy are submitted to an open access register before beginning the study.
Randomisation	2. Each participant is randomly assigned to the experimental or control group.
	3. In quasi-experimental studies, groups of participants are randomly assigned to the experimental or control condition.
Blinding	4. Participants do not know if they belong to the experimental or to the control group.
	5. Instructors do not know if they belong to the experimental or to the control group.
	6. People in charge of evaluating the results do not know if they belong to the experimental or to the control group.
Similarities between groups	7. Socio-economic characteristics of experimental and control groups are proven to be similar.
	8. Pre-test scores on the dependent variable(s) of interest are proven to be similar in the experimental and control group.
Training of instructor	9. The instructor receives training and development in MIT applied.
Intervention fidelity	10. The study does not include "non-valid" practices on the basis of Gardner's guidelines (1995).
Type of control group	11. The study includes an active control group (as opposed to a passive control group).
Information for replicability	12. The information provided is enough to replicate the intervention.
	13. The information about the tool employed to measure the dependent variable is fully reported.
Reliability of tests	14. The study offers a satisfactory reliability rating of the tests employed to measure the dependent variable(s).
Validity of tests	15. The study offers a satisfactory validity rating of the tests employed to measure the dependent variable(s).
Availability of statistic data	
	16. Results of at least one key comparison between experimental and control group are reported.
Open access publication	17. The study is freely available on-line.

research. To mention just some noticeable examples, Item 4, referring to the blinding of participants, is only explicitly described in five studies (12.82%), while Item 15, related to the validity of the tests employed to measure the dependent variable, is not explicitly addressed in any study. In addition, most studies fail to meet a substantial number of quality criteria. For example, Item 11, referring to the use of an active control group, is only fulfilled by one study (2.56%), while the remaining studies either use a passive control group that is trained with a traditional method (87.18%) or do not provide any information on this issue (10.26%). And Item 12, assessing whether the articles report sufficient information to replicate the intervention, is only addressed in three studies (7.69%). Of particular interest is Item 10, related to intervention fidelity, where the majority of studies (69.23%) do not give any information about the activities included in the intervention. Six studies (15.38%) reported well–described proposals but some of them belong to the categories labelled as *not right* by Gardner himself. Only 6 (15.38%) studies fulfilled both conditions, that is, reporting sufficient information and comprising sound activities, according to Gardner's criteria. Just four quality criteria were fulfilled by more than 50% of the studies: Item 7, referring to the equal socioeconomic variables between groups; Item 8, referring to the analysis of the pre-test scores; Item 14, referring to the reliability of the tests employed to measure the dependent variable, and Item 16, related to the availability of information about at least one of the key comparisons reported in the analysis of results.

3.3. *Quantitative meta–analysis*

Among the 26 standardized mean change difference scores entered in the meta–analysis, the average mean effect was remarkably large, $g_\Delta = 1.68$, 95% CI [1.13, 2.22], and statistically significant, $z = 6.01$, $p < .001$. The level of heterogeneity was also large and significant, $I^2 = 97.70\%$, $Q(25) = 498.50$, $p < .001$. Similarly, among the 28 standardized mean differences scores that could be entered in the meta–analysis of post–test measures, the mean average effect size was very large and statistically significant, $g_p = 1.25$, 95% CI [0.93, 1.57], $z = 7.72$, $p < .001$, and the distribution of effect sizes was highly heterogeneous, $I^2 = 91.25\%$, $Q(27) = 230.21$, $p < .001$.

The suite of outlier-detection methods implemented in the influence function of the 'metafor' package allowed us to identify an outlier in the first meta-analysis (with a $g_\Delta = 7.16$) and another one in the second one (with a $g_p = 3.76$). After removing these outliers, the average effect sizes declined slightly, $g_\Delta = 1.49$, 95% CI [1.05, 1.93], $z = 6.66$, $p < .001$, and and $g_p = 1.15$, 95% CI [0.88, 1.42], $z = 8.27$, $p < .001$, and so did heterogeneity, $I^2 = 96.42\%$, $Q(24) = 437.70$, $p < .001$, and $I^2 = 87.76\%$, $Q(26) = 173.78$, $p < .001$, respectively. Figs. 4 and 5 show the point estimates and confidence intervals of all the studies included in these two meta-analyses, together with their meta-analytic averages. As can be seen, even after removing outliers, both types of effect sizes span a wide range of values, including a substantial number of effect sizes with extremely large values.

The funnel plots for both types of effect size, depicted in Fig. 6, reflect that the distribution of effect sizes with respect to their standard errors is not symmetric. In general, the largest effect sizes come from studies with the largest standard errors (i.e. with the smallest samples). The red regression lines depicted in both funnel plots are statistically significant, $b_1 = 6.79$, $z = 4.57$, $p \leq 0.001$, for g_Δ, and $b_1 = 5.05$, $z = 3.99$, $p < .001$, for g_p. With some caveats (Sterne et al., 2011), this asymmetric distribution is usually taken as an indicator of publication bias or selective reporting and, in any case, suggests that the results of the meta–analysis should be interpreted with extreme caution because the reported effect sizes may overestimate the size of the true effects (Egger, Davey Smith, Schneider, & Minder, 1997).

To further explore publication bias and selective reporting, we applied four different methods for the detection and correction of bias: PET, PEESE, trim-and-fill, and Vevea and Hedges' (1995) selection model. PET (Stanley & Doucouliagos, 2013) is based on a regression of effect sizes on standard errors, like the ones depicted in Fig. 6. Using this regression, it is possible to predict what would be the expected effect size for an ideal experiment with zero standard error. PEESE (Stanley & Doucouliagos, 2013) is based on the same logic, with the only exception that the regression predicts effects sizes from study variances instead of the standard errors. Trim-and-fill (Duval & Tweedie, 2000) identifies which studies contribute most to funnel plot asymmetry and then restores asymmetry by inputing an equal number of missing studies on the other side of the funnel plot. Finally, Vevea and Hedges' selection model assumes that studies with non-significant results might be underrepresented in the meta-analysis and corrects the average effect size after estimating the likely magnitude of the bias against non-significant results.

When applied to our data sets, PET detected significant signs of bias for both types of effect size, $z = 4.57$, $p < .001$, for g_Δ and $z = 3.99$, $p < .001$, for g_p. The bias corrected estimates were $g_\Delta = -0.28$, 95% CI [−1.09, 0.53], and $g_p = -0.26$, 95% CI [−0.97, 0.46], respectively. PEESE also detected significant bias for g_Δ, $z = 2.25$, $p = .002$, and for g_p, $z = 3.67$, $p < .001$, and yielded bias-corrected estimates of $g_\Delta = 0.93$, 95% CI [0.42, 1.44], and $g_p = 0.60$, 95% CI [0.24, 0.96]. Trim-and-fill estimated that the number of missing studies was 10 for g_Δ and 2 for g_p. The bias-corrected estimates were 0.71, 95% CI [0.17, 1.25], and 1.06, 95% CI [0.76, 1.35], respectively. Finally, the selection model did not detect evidence of bias in either g_Δ, $\chi^2(1) = 2.49$, $p = .114$, or g_p, $\chi^2(1) = 1.70$, $p = .193$, and yielded bias-corrected estimates of 1.02, 95% CI [0.04, 1.99], and 0.94, 95% CI [0.43, 1.45], respectively. In sum, three out of the four tests suggest that the average estimates of our meta-analyses could be inflated by publication or reporting biases. One of

	1. Pre-registration	2. Participants are randomised	3. Groups are randomised	4. Participants are blinded	5. Instructors are blinded	6. Analysts are blinded	7. Groups are equal in socio-economic variables	8. Pre-test scores are analysed	9. Instructors receive training	10. Intervention fidelity is preserved	11. Active control group	12. Enough information to replicate the intervention	13. Enough information to replicate the DV	14. Satisfactory reliability of tests employed	15. Satisfactory validity of tests employed	16. Results of at least one key comparison reported	17. Open data access
Abdi et al. (2013)	-	-	-	?	?	?	+	+	+	?	-	-	-	+	?	+	-
Abdulkader et al. (2009)	-	+	+	-	-	?	+	+	+	-	?	-	-	+	?	+	-
Akkuzu et al. (2011)	-	?	?	?	-	?	?	-	+	-	-	-	-	+	?	-	-
Alavinia et al. (2012)	-	-	+	?	?	-	?	-	?	?	-	-	-	?	?	-	-
Al-Balhan (2006)	-	+	+	?	-	?	?	+	+	?	-	-	-	?	?	+	-
Alqatanani (2017)	-	-	+	?	?	-	+	-	?	+	-	-	-	+	?	+	-
Altintas et al. (2015)	-	-	?	?	?	?	?	+	?	?	?	-	-	+	?	+	-
Al-Zoubi et al. (2019)	-	+	+	?	-	?	+	+	+	?	-	-	-	+	?	+	-
Anaduaka (2008)	-	-	+	?	?	-	+	-	+	?	-	-	-	+	?	+	-
Bilgin (2006)	-	-	+	?	?	?	+	+	+	?	-	+	+	-	?	+	-
Delgoshaei et al. (2012)	-	-	-	?	?	?	+	-	?	?	-	-	-	+	?	+	-
Dillihunt et al. (2006)	-	-	+	?	-	-	?	-	+	+	-	-	-	+	?	+	-
Eng et al. (2010)	-	?	?	?	?	+	+	-	?	+	-	-	-	?	?	+	-
Fasni et al. (2017)	-	-	-	?	?	?	?	+	?	?	?	-	-	?	?	+	-
Gün et al. (2014)	-	-	+	?	-	?	+	+	+	?	-	-	-	+	?	+	-
Gündüz et al. (2016)	-	+	+	?	-	+	+	+	+	?	-	-	-	+	?	+	-
Gurbuz et al. (2014)	-	-	+	?	-	?	+	+	+	+	-	-	-	+	?	+	-
Gurcay et al. (2017)	-	-	+	-	?	?	?	-	?	+	-	-	-	+	?	+	-
Haboush (2010)	-	?	?	?	-	?	+	+	+	?	-	-	+	+	?	+	-
Hanley et al. (2002)	-	?	?	?	-	?	+	-	+	?	-	-	+	-	-	-	-
Inan et al. (2017)	-	?	?	?	?	?	+	+	?	?	-	-	-	+	?	-	-
Işık et al. (2009)	-	?	?	-	-	?	+	+	+	?	-	-	-	+	?	+	-
Kaya et al. (2007)	-	+	+	?	?	?	+	-	+	?	-	-	-	+	?	+	-
Khalghollah et al. (2014)	-	+	+	?	?	?	+	-	?	?	-	-	-	?	?	+	-
Liliawati et al. (2018)	-	-	-	?	?	?	+	-	?	?	-	-	-	-	?	-	-
Modirkhamene et al. (2012)	-	-	?	?	?	?	?	+	?	-	+	-	-	+	?	+	-
Mostafa (2017)	-	+	+	?	-	?	?	+	+	-	?	-	-	+	?	+	-
Nakhbi et al. (2016)	-	-	?	?	-	?	?	-	?	?	-	-	-	-	?	-	-
Nuallaong et al. (2015)	-	-	-	?	?	?	+	+	?	?	-	-	-	+	?	+	-
Özdermir et al. (2006)	-	-	+	?	-	?	+	+	?	-	-	-	-	+	?	+	-
Pahlavani et al. (2017)	-	-	-	-	-	?	?	-	?	?	-	-	+	+	?	+	-
Safranj et al. (2018)	-	?	?	?	?	?	?	-	?	+	-	-	-	?	?	+	-
Sanchez-Martin et al. (2017)	-	?	?	?	?	?	+	+	?	?	-	-	-	?	?	+	-
Sheahan et al. (2015)	-	+	+	-	-	?	+	+	+	?	-	+	-	-	-	+	-
Soleimani et al. (2012)	-	+	+	?	?	?	+	-	+	?	-	-	-	+	?	+	-
Stanciu et al. (2011)	-	?	?	?	?	?	?	+	?	?	-	-	-	?	?	+	-
Ucak et al. (2006)	-	-	-	?	?	?	+	-	?	?	-	-	-	+	?	-	-
Yalmanci et al. (2013)	-	-	-	?	-	?	?	+	?	-	-	-	-	+	?	-	-
Young (2003)	-	-	+	?	?	?	+	+	?	?	-	+	+	?	?	+	-

Fig. 2. Quality assessment and values assigned to each item.

them, PET, suggests that the bias-corrected averages might be non-significant. PEESE and trim-and-fill, in contrast, suggest that the true effect might be different from zero although substantially smaller than suggested by the uncorrected meta-analytic estimates.

4. Discussion

Since Gardner developed his theory about the existence of multiple intelligences, a growing number of teachers have adapted and incorporated the theory into their daily routine in the classroom (White, 2004). In spite of this unexpected success, as Gardner himself has

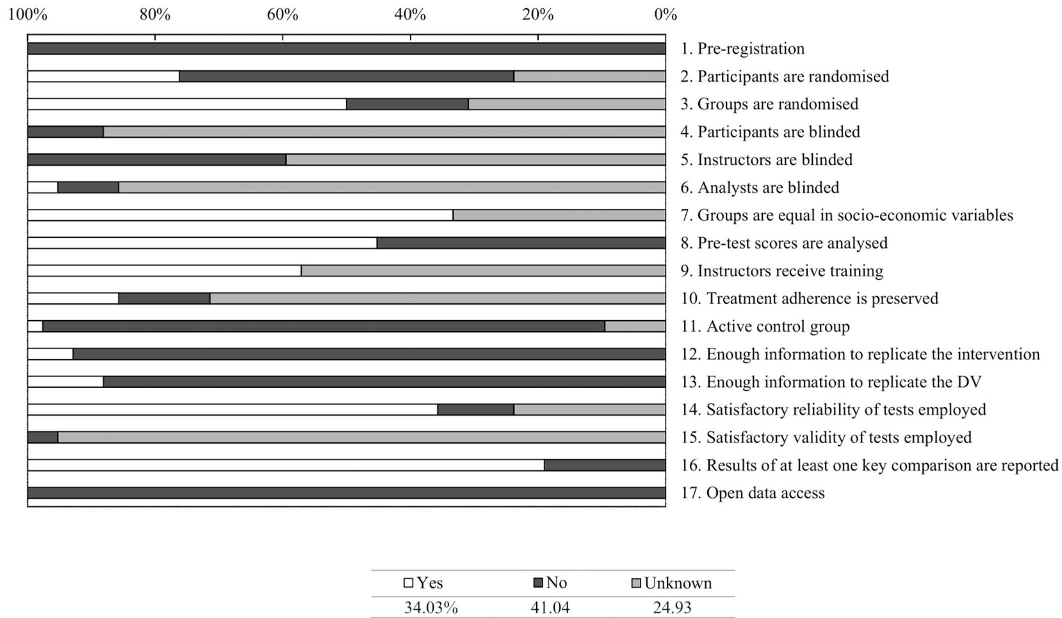

Fig. 3. Graphical representation of the values obtained in the scale of quality.

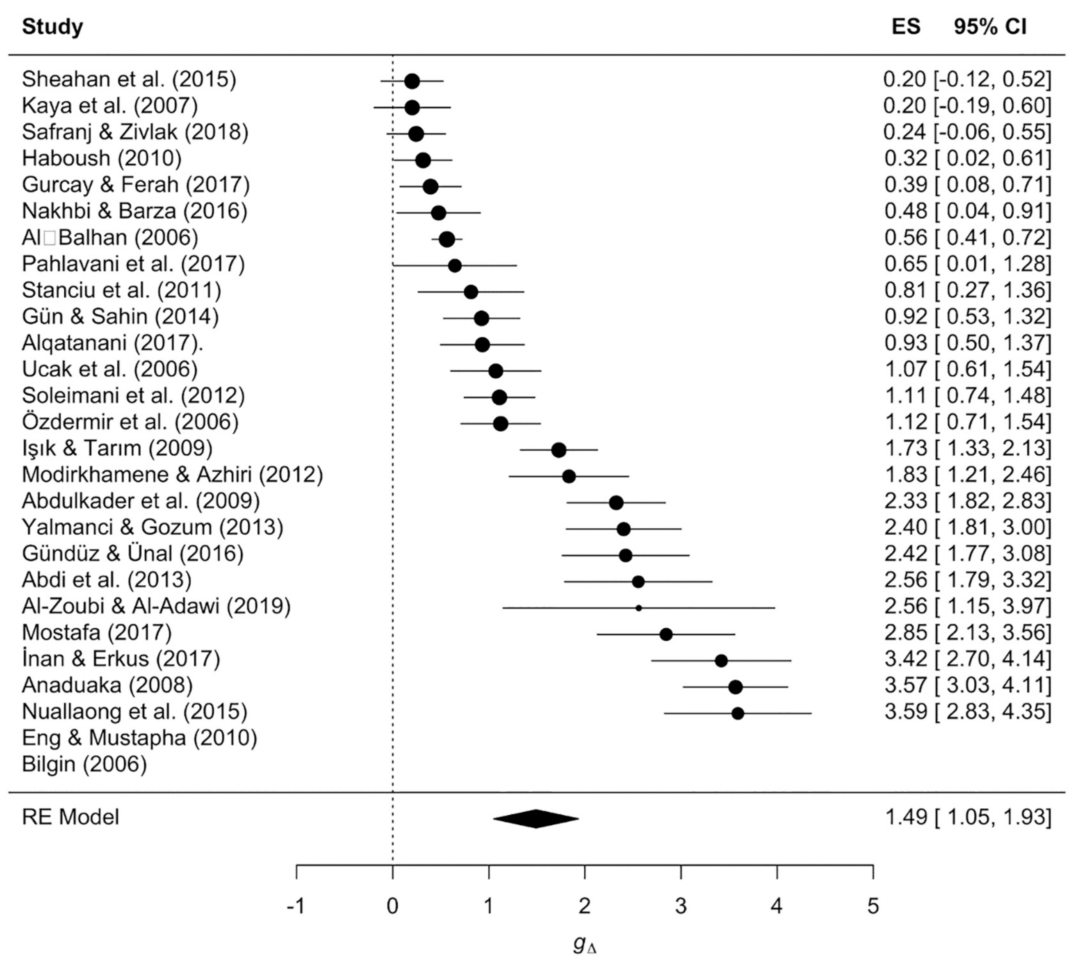

Fig. 4. Forest plot for the meta-analysis of g_Δ scores.

recurrently recognized, there is no solid data about the effectiveness of applying MIT–inspired interventions in the academic achievement of students. To date there are only two meta–analyses on this matter and, as we have discussed above, both of them present important methodological shortcomings, such as an absence of any assessment of the quality of the studies included or a lack of control for publication bias. The aim

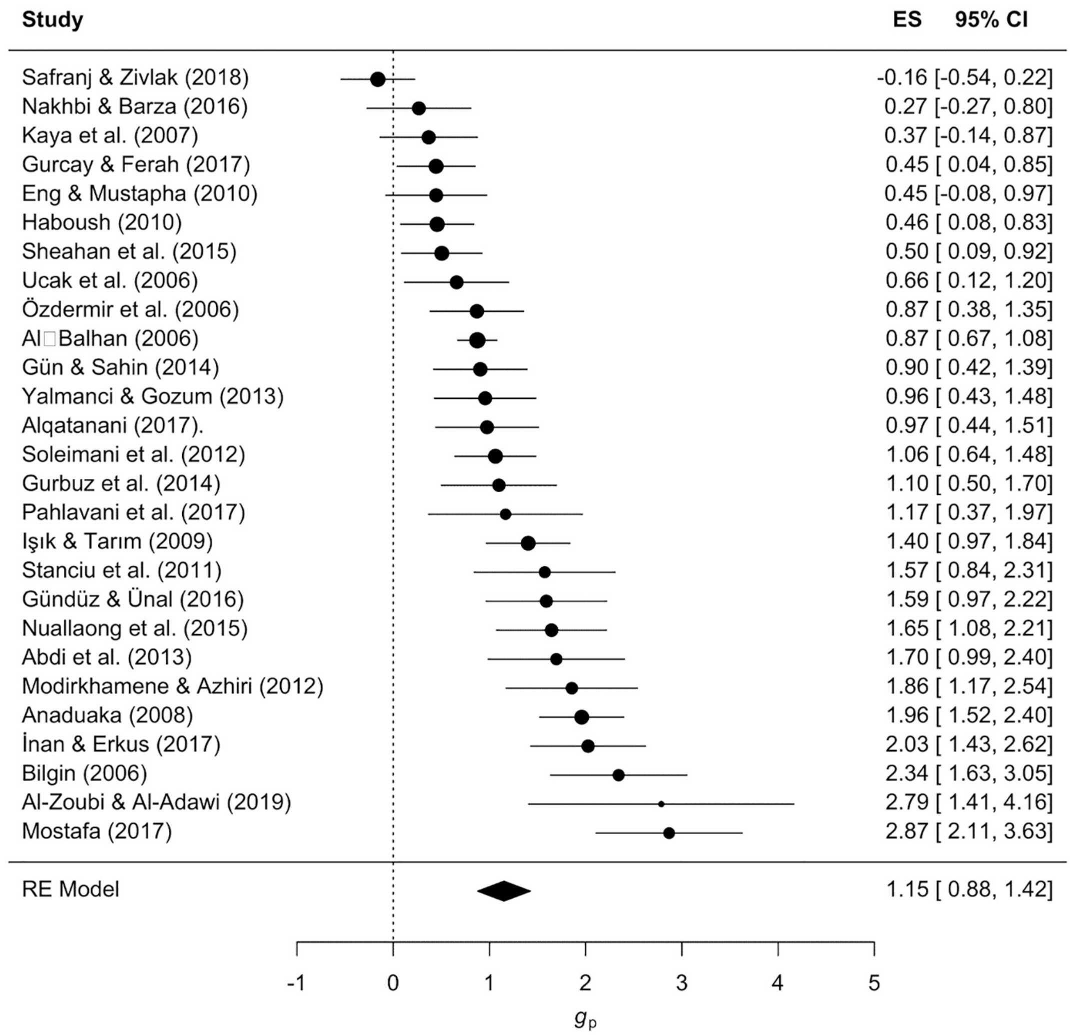

Fig. 5. Forest plot for the meta-analysis of g_p scores.

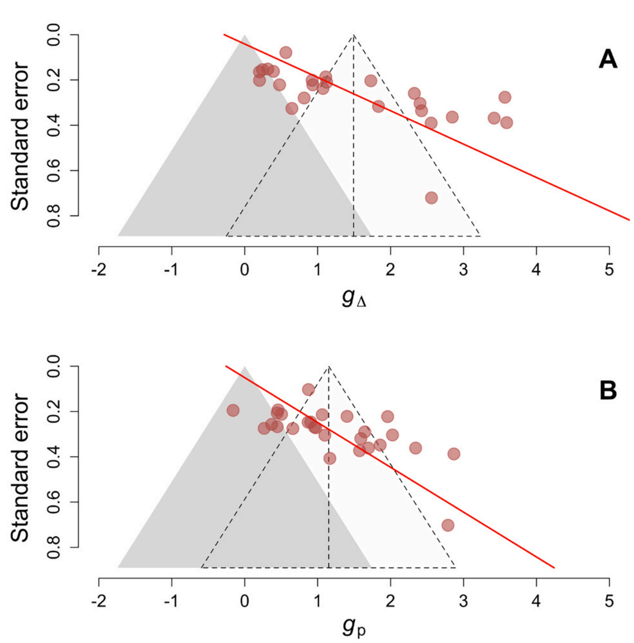

Fig. 6. Funnel plots for the meta-analysis of g_Δ scores (panel A) and g_p scores (panel B).

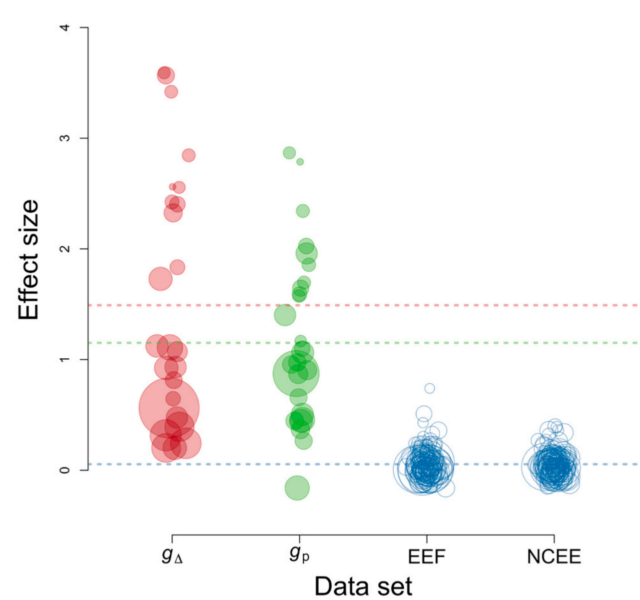

Fig. 7. Distribution of effect sizes in the present meta-analysis (g_Δ and g_p) and two data sets of pre-registered studies (EEF and NCEE).

of the present systematic review was to assess the quality of the studies testing the impact of MIT–inspired instructional methodologies on academic achievement of learners, overcoming the existing flaws of previous reviews as much as possible.

In general, the qualitative analysis of the results showed that the studies included in this review have important methodological flaws and report insufficient information about essential elements to make a critical appraisal of the methods, such as whether participants and instructors were blind to experimental manipulation, or whether the measures employed were reliable and valid. Perhaps more importantly, only a handful of studies described the intervention undertaken in sufficient detail to allow its replication. In other words, there is no way of knowing what the interventions consisted of and how the dependent variable was measured. When methodological information was given, many of the studies failed to meet important quality criteria, such as the randomisation of participants or the inclusion of an active control group. In fact, only a couple of quality criteria were clearly fulfilled by the majority of studies.

The quantitative analysis of the data replicates the results of previous meta–analyses, but with important caveats. As explained in the introduction, Bas (2016) and Batdi (2017) reported large effect sizes for MIT–based interventions ($d = 1.077$ and 0.95, respectively). Consistent with them, we find remarkably large effect sizes of $g_\Delta = 1.49$ and $g_p = 1.15$. The sheer size of these effects should, on its own, be sufficient reason for skepticism (Pashler, Rohrer, Abramson, Wolfson, & Harris, 2016). To put these effect sizes in proper context, Fig. 7 shows the distribution of g_Δ and g_p from the studies included in the present meta–analysis, together with the effect sizes (standardized mean differences) of two large sets of high–quality educational studies commissioned by the Education Endowment Foundation (EEF) in the UK and the National Center for Educational Evaluation and Regional Assistance (NCEE) in the USA (Lortie-Forgues & Inglis, 2019). It is clear that the effects reported for the MIT–based interventions reviewed here are remarkably larger than the effects reported by the studies funded by these two institutions. They are also much larger than the typical effect sizes reported in psychological research (Funder & Ozer, 2019; Rubio-Aparicio, Marín-Martínez, Sánchez-Meca, & López-López, 2018).

What factors could explain the striking difference between the effect sizes found in the present studies and those reported in other areas of educational research? The funnel plots depicted in Fig. 6 offer a plausible response to this question. As can be seen, the largest effect sizes come from the studies with the lowest precision, that is, with the smallest number of participants. This pattern of results suggests that the average effect size is probably inflated by the (large) results of the lowest–quality studies.

In addition, all the studies commissioned by the EEF and the NCEE are required to meet the highest methodological standards, including the use of powerful sample sizes, active control groups, reliable and valid outcome measures, preregistered methods and analyses, and unconditional publication regardless of outcome (Lortie-Forgues & Inglis, 2019). In comparison, Fig. 2 shows that only a handful of the studies reviewed here complied with these standards. Only one of the studies included an active control group. This is unfortunate, because the available evidence shows that educational studies relying on passive control groups yield grossly overestimated effect sizes (Sala & Gobet, 2017). In fact, the inclusion of an active control group has been considered a decisive measure to test the efficacy of educational interventions (e.g. Datta, 2007), as long as the expectations of students in an active control group is guaranteed to be the same as the ones of those in the experimental group (Boot, Simons, Stothart, & Stutts, 2013).

None of the studies were preregistered, which, again, is an essential protection against biases in research (Kaplan & Irvin, 2015; Warren, 2018) as it reduces researchers' degree of freedom and questionable research practices, such as the selective publication of analyses that "worked" (Simmons, Nelson, & Simonsohn, 2011). Similarly, measurement error can inflate effect sizes when a population effect size is estimated across small sample sizes (Loken & Gelman, 2017), a bias whose impact on the present studies is difficult to estimate because most of them failed to report psychometric information about the dependent measures. Fig. 2 also shows that none of the articles reviewed explicitly stated that participants and instructors were blind to experimental manipulation, which means that the results of the interventions could be entirely due to the positive expectations of participants, as mentioned above (Boot et al., 2013). Although difficult, it is possible to blind participants and instructors through the use of active control groups where the actors involved do not know whether they are being trained by the intervention under study or under an alternative one.

Given these caveats (and other problems highlighted in Fig. 2), the fact that the effect sizes reported in this literature are large is unsurprising. In our opinion, this literature should not be taken as evidence that MIT–based interventions work. All in all, although the majority of studies included in the present work suggested that MIT–inspired interventions yielded significant improvements in the academic achievement of students, it is imperative to interpret these results in the light of critical shortcomings that have emerged in the qualitative and quantitative analyses of the data.

To put these results in context, it is also important to note that the main tenet of MIT about the existence of multiple intelligences is not supported by the scientific community. Research in cognitive psychology has systematically pointed out the existence of a single intelligence, or general factor, that explains most of the variance in cognitive performance in different tasks (Lubinski, 2004; Visser et al., 2006a). Most relevant for this study, the central claim regarding the application of MIT in schools lacks sound evidence. Presumably, all the intelligences should be used as channels when presenting new materials so that students experience the material via their best intelligence, and thus understanding will be promoted. However, studies in the field of learning psychology have shown that the best way to learn something is usually defined by the content itself, and not by the particular abilities or, in terms of Gardner, the specific intelligences profile of learners (Willingham, 2004). In other words, according to the best evidence available so far, teaching should be subordinated to the object of learning, not to the characteristics of individual learners.

Aside from these important gaps in the theory and its translation into classroom practice, any attempt to test the efficacy of MIT–inspired interventions in the future should address the methodological flaws of the existing literature that we have highlighted in the present review. Ideally, these studies should adopt experimental designs, use large samples, guarantee the blinding of participants and instructors, include an active control group, and follow detailed reporting guidelines, including precise information about the sample, procedure and materials employed in study, so that the results can be replicated by independent researchers.

MIT might have contributed to rethinking some important questions among educators, such as the fact that children are unique and valuable regardless of their capacities and that schools are responsible for helping all of them bring out their best and find their real interests and strengths. Or the fact that, too often, schools have exclusively focused on purely academic skills, such as reading or mathematics, at the expense of other skills, such as music or corporal expression, leading many children to fail in finding their real interests and strengths. Bearing this undeniable contribution to education in mind, it is understandable that many teachers have embraced MIT-inspired interventions in the classroom with great enthusiasm. However, as shown in the present study, the evidence gathered to date on the effectiveness of these educational actions does not allow for a valid assessment of their impact on learning. Due to the importance of implementing class well-grounded methods of instruction (Cook & Cook, 2004), it is imperative to perform high-quality research on the effectiveness of MIT-based intervention before its use in the classroom can be recommended or promoted.

Acknowledgements

MF was supported by bizkaia:talent, Diputación Foral de Bizkaia [grant number AYD-000-235]; and by Programa Posdoctoral de Perfeccionamiento de Personal Investigador Doctor, Gobierno Vasco. MAV was supported by Comunidad de Madrid [grant number 2016-T1/SOC-1395]; and AEI/FEDER UE [grant number PSI2017-85159-P.

References

Abdi, A., Laei, S., & Ahmadyan, H. (2013). The effect of teaching strategy based on multiple intelligences on students' academic achievement in science course. *Universal Journal of Educational Research, 1*, 281–284.

Abdulkader, F. A., Gundogdu, K., & Mourad, E. A. (2009). The effectiveness of a multiple intelligences–based program on improving certain reading skills in 5th–year primary learning disabled students. *Electronic Journal of Research in Educational Psychology, 71*, 673–690. https://doi.org/10.25115/ejrep.v7i18.1340.

Akkuzu, N., & Akçay, H. (2011). The design of a learning environment based on the theory of multiple intelligence and the study its effectiveness on the achievements, attitudes and retention of students. *Procedia Computer Science, 3*, 1003–1008. https://doi.org/10.1016/j.procs.2010.12.165.

Alavinia, P., & Farhady, S. (2012). Using differentiated instruction to teach vocabulary in mixed ability classes with a focus on multiple intelligences and learning styles. *International Journal of Applied Science and Technology, 2*, 72–79.

Al-Balhan, E. M. (2006). Multiple intelligence styles in relation to improved academic performance in Kuwaiti middle school reading. *Digest of Middle East Studies, 15*, 18–34. https://doi.org/10.1111/j.1949-3606.2006.tb00002.x.

Alqatanani, A. (2017). Do multiple intelligences improve EFL students critical reading skills? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2945910.

Altıntaş, E., & Özdemir, A. S. (2015). The effect of the developed differentiation approach on the achievements of the students. *Eurasian Journal of Educational Research, 15*, 199–216. https://doi.org/10.14689/ejer.2015.61.11.

Al-ZoubiI, S., & Al-Adawi, F. A. (2019). Effects of instructional activities based on multiple intelligences theory on academic achievement of Omani students with dyscalculia. *Journal for the Education of Gifted Young Scientists, 7*, 1–14.

Amstrong, T. (2009). *Multiple intelligences in the classroom* (3rd ed.). USA: ASCD.

Anaduaka, U. S. (2008). *Effect of multiple intelligences teacher approach on students achievement and interest in geometry*. University of Nigeria, Nssuka, Nigeria: Doctoral dissertation. Retrieved from https://oer.unn.edu.ng/read/effect-of-multiple-intelligences-teacher-approach-on-students-achievement-and-interest-in-geometry-4?rdr=1.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*, 3–25.

Armstrong, T. (1994). *Multiple intelligences in the classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Bas, G. (2016). The effect of multiple intelligences theory–based education on academic achievement: A meta–analytic review. *Educational Sciences: Theory and Practice, 16, 183–1864*. https://doi.org/10.12738/estp.2016.6.0015.

Batdi, V. (2017). The effect of multiple intelligences on academic achievement: A meta–analytic and thematic study. *Educational Sciences: Theory and Practice, 17*, 2057–2092. https://doi.org/10.12738/estp.2017.6.0104.

Bilgin, E. K. (2006). *The effect of multiple intelligences based instruction on ninth graders chemistry achievement and attitudes toward chemistry*. Middle East Technical University, Ankara, Turkey: Doctoral dissertation. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.633.7329&rep=rep1&type=pdf.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem of with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*, 445–454. https://doi.org/10.1177/1745691613491271.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta–analysis*. Chichester, United Kingdom: John Wiley & Sons. https://doi.org/10.1002/9780470743386.

Candler, L. (2011). Multiple Intelligence Survey for Kids. Retrieved from: http://www.lauracandler.com/free/misurvey.

Chan, D. W. (2006). Perceived multiple intelligences among male and female Chinese gifted students in Hong Kong: The structure of the student multiple intelligences profile. *National Association for Gifted Children, 50*, 325–338.

Chen, J. Q., Feldman, D. H., Gardner, H., & Krechevsky, M. (1998a). In *Project Zero framework for early learning. Volume 1, building on children's strengths. The experience of project Spectrum*. New York, NY: Teacher College Press (Gen. Eds.).

Chen, J. Q., Feldman, D. H., Gardner, H., & Krechevsky, M. (1998b). In *Project Zero framework for early learning. Volume 2, project Spectrum. Early learning activities*. New York, NY: Teacher College Press (Gen. Eds.).

Chen, J. Q., Feldman, D. H., Gardner, H., & Krechevsky, M. (1998c). In *Project Zero framework for early learning. Volume 3, project Spectrum. Preschool assessment handbook*. New York, NY: Teacher College Press (Gen. Eds.).

Chen, J. Q., Moran, S., & Gardner, H. (2009). *Multiple intelligences theory around the world*. San Francisco: Jossey–Bass.

Collin, J. (2001). Seven kind of smart. *Time.*, (June 24) Retrieved from: http://content.time.com/time/magazine/article/0,9171,140234,00.html.

Colom, R. (2018). *Manual de psicología diferencia. Métodos modelos y aplicaciones*. Madrid, Spain: Pirámide.

Cook, B. G., & Cook, L. (2004). Bringing science into the classroom by basing craft on research. *Journal of Learning Disabilities, 37*, 240–247.

Datta, L. E. (2007). Why an active control group makes a difference and what to do about it. *Journal of MultiDisciplinary Evaluation, 4*, 1–12.

Delgoshaei, Y., & Delavari, N. (2012). Applying multiple–intelligence approach to education and analyzing its impact on cognitive development of pre–school children. *Procedia - Social and Behavioral Sciences, 32*, 361–366.

Dillihunt, M. L., & Tyler, K. M. (2006). Examining the effects of multiple intelligence instruction on math performance. *Journal of Urban Learning, Teaching, and Research, 2*, 131–150.

Duval, S., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *The British Medical Journal, 315*, 629–634.

Eng, L. L., & Mustapha, G. (2010). Enhancing writing ability through multiple–intelligence strategies. *Pertanika Journal of Social Sciences and Humanities, 18*, 53–63.

Fasni, N., Fatimah, S., & Yulanda, S. (2017, May). The implementation of multiple intelligences based teaching model to improve mathematical problem solving ability for student of junior high school. In *, Vol. 1848*. *AIP Conference Proceedings* (p. 040011). AIP Publishing. https://doi.org/10.1063/1.4983949. No. 1.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*, 156–168. https://doi.org/10.1177/2515245919847202.

Gardner, H. (1983). *Frames of mind. The theory of multiple intelligences*. New York, NY: Basic Books.

Gardner, H. (1991). *The unschooled mind: How children think and how schools should teach*. New York, NY: Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York, NY: Basic Books.

Gardner, H. (1995). Reflections on multiple intelligences: Myths and messages. *Kappa, 77*, 202–209.

Gardner, H. (1997). Multiple intelligences as a partner in school improvement. *Educational Leadership*, 20–21.

Gardner, H. (2006). On failing to grasp the core of MI theory: A response to Visser et al. *Intelligence, 34*, 503–505. https://doi.org/10.1016/j.intell.2006.04.002.

Gardner, H. (2011). *The unschooled mind. How children think and how schools should teach* (20th ed.). New York: Basic Books.

Gardner, H., & Hatch, H. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher, 18*, 4–10. https://doi.org/10.2307/1176460.

Gardner, H., & Morgan, S. (2006). The science of multiple intelligences theory: A response to Lynn Waterhouse. *Educational Psychologist, 41*, 227–232. https://doi.org/10.1207/s153226985ep4104_2.

Geake, J. (2008). Neuromythologies in education. *Educational Research, 50*, 123–133. https://doi.org/10.1080/00131880802082518.

Giorgis, N. (2007). *Multiple Intelligences Profiles. Electronic Bulletin Engineering First, 5*.

Gün, E. S., & Sahin, A. E. (2014). The effect of layered curriculum supported by multiple intelligences on students' achievements and permanence. *Egitim ve Bilim, 39*. https://doi.org/10.15390/eb.2014.2424.

Gündüz, Z. E., & Üna, I. D. (2016). Effects of multiple intelligences activities on writing skill development in an efl context. *Universal Journal of Educational Research, 4*, 1687–1697.

Gurbuz, R., Birgin, O., & Catlioglu, H. (2014). The effects of activities based on the multiple intelligence theory of students' conceptual learning and their retention: A case of circle and cylinder. *Online Submission, 9*, 197–205.

Gurcay, D., & Ferah, H. O. (2017). The effects of multiple intelligences based instruction on students' physics achievement and attitudes. *Journal of Baltic Science Education, 16*, 666–677.

Haboush, Z. Y. (2010). *The effectiveness of using a programme based on multiple intelligences theory on eighth graders' English reading comprehension skills*. The Islamic University of Gaza, Gaza, Palestine: Doctoral dissertation. Retrieved from https://iugspace.iugaza.edu.ps/handle/20.500.12358/17593.

Hanley, C., Hermiz, C., Lagioia-Peddy, J., & Levine-Albuck, V. (2002). *Improving student interest and achievement in social studies using a multiple intelligence approach (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (ED No. 465696)*.

Harvard Graduate School of Education. (2016). Project Zero. Retrieved from: http://www.pz.harvard.edu/projects/.

Harzing, A. W. (2007). Publish or Perish. available from https://harzing.com/resources/publish-or-perish.

Hunt, E. B. (2001). *Human intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511781308.

İnan, C., & Erkus, S. (2017). The effect of mathematical worksheets based on multiple intelligences theory on the academic achievement of the students in the 4th grade primary school. *Universal Journal of Educational Research, 5*, 1372–1377.

Işık, D., & Tarım, K. (2009). The effects of the cooperative learning method supported by multiple intelligence theory on Turkish elementary students' mathematics achievement. *Asia Pacific Education Review, 10*, 465. https://doi.org/10.1007/s12564-009-9049-5.

Jensen, A. R. (1998). The g factor: The science of mental ability. In *Human evolution, behavior, and intelligence*. Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA, 282*, 1054–1060.

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects on large NHLBI clinical trials has increased over time. *PLoS One, 10*, Article e0132382. https://doi.org/10.1371/journal.pone.0132382.

Kaya, O. N., Dogan, A., Gokcek, N., Kilic, Z., & Kilic, E. (2007). *Comparing multiple intelligences approach with traditional teaching on eight grade students' achievement in and attitudes toward science*. Chicago, IL: The Annual Meeting of the American Educational Research Association (April 8-13).

Khalghollah, M., Afsha, J., & Shahidi, N. (2014). Comparison of the effect of based on multiple intelligences and methods of learning science 7–8 year female students primary schools in Shiraz in the 2013–2014 years. *International Journal of Biology, Pharmacy and Allied Sciences, 3*, 2875–2880.

Liliawati, W., Purwanto, Zulfikar, A., & Kamal, R. N. (2018). The effectiveness of learning materials based on multiple intelligence on the understanding of global warming. *Journal of Physics: Conference Series, 1013*, Article 012049. https://doi.org/10.1088/1742-6596/1013/1/012049.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science, 355*, 584–585. https://doi.org/10.1126/science.aam5409.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large–scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher, 48*, 158–166. https://doi.org/10.3102/0013189X19832850.

Lubinski, D. (2004). Introduction to the special section of cognitive abilities:100 years after Sperman's (1904), "general intelligence", objectively determined and measured. *Journal of Personality and Social Psychology, 86*, 96–111. https://doi.org/10.1037/0022-3514.86.1.96.

McKenzie, W. (1999). Multiple Intelligences Survey. Retrieved from http://surfaquarium.com/MI/inventory.htm.

Modirkhamene, S., & Azhiri, M. H. B. (2012). The effect of multiple intelligences–based reading tasks on EFL learners' reading comprehension. *Theory and Practice in Language Studies, 2*, 1013. https://doi.org/10.4304/tpls.2.5.1013-1021.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*, b2535. https://doi.org/10.1136/bmj.b2535.

Morris, S. B. (2008). Estimating effect sizes from pretest–posttest–control group designs. *Organizational Research Methods, 11*, 364–386. https://doi.org/10.1177/1094428106291059.

Mostafa, A. A. (2017). The effect of using multiple intelligences on some basic reading skills of first graders at–risk for reading disabilities. *International Journal of Psycho–Educational Sciences, 6*, 109–116.

Nakhbi, K. A., & Barza, L. (2016). The effectiveness of using multiple intelligence strategy on sixth grade students' achievements and attitudes toward science. *Australian Journal of Sustainable Business and Society, 2*, 29–37. https://doi.org/10.3926/jotse.404.

Newman, M. (2003). A pilot systematic review and meta–analysis on the effectiveness of problem–based learning. In *On behalf of the Campbell Collaboration Systematic Review Group on the effectiveness of problem–based learning*. Newcastle, UK: University of Newcastle, Learning and Teaching Support Network.

Nofal, M. (2010). *Multiple intelligence in the classroom*. Jordan: AmmanMassira for Publishing.

Nuallaong, W., Nuallaong, T., & Preechadirek, N. (2015). Academic achievement from using the learning medium via a tablet device based on multiple intelligences in grade 1 elementary student. *Journal of the Medical Association of Thailand, 98*, S24–S28.

Özdermir, P. I., Güneysu, S., & Tekkaya, C. (2006). Enhancing learning through multiple intelligences. *Journal of Biological Education, 40*, 74–78. https://doi.org/10.1080/00219266.2006.9656017.

Pahlavani, A., Khosravani, E., & Zanjani, F. (2017). On the effects of linguistic intelligence–based activities on Iranian EFL learners' speaking ability. *Modern Journal of Language Teaching Methods, 7*, 144–150.

Pashler, H., Rohrer, D., Abramson, I., Wolfson, T., & Harris, C. R. (2016). A social priming data set with troubling oddities. *Basic and Applied Social Psychology, 38*, 3–18. https://doi.org/10.1080/01973533.2015.1124767.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Beverly Hills, CA: Sage.

Rubio-Aparicio, M., Marín-Martínez, F., Sánchez-Meca, J., & López-López, J. A. (2018). A methodological review of meta–analyses of the effectiveness of clinical psychology treatments. *Behavior Research Methods, 50*, 2057–2073. https://doi.org/10.3758/s13428-017-0973-8.

Saban, A. (2005). *Çoklu Zeka Teorisi ve Eğitim*. Ankara: Nobel Yayınları.

Safranj, J., & Zivlak, J. (2018). Spatial–visual intelligence in teaching students of engineering. *Research in Pedagogy, 8*, 71. https://doi.org/10.17810/2015.72.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science, 26*, 515–520. https://doi.org/10.1177/0963721417712760.

Sánchez-Martín, J., Álvarez-Gragera, G. J., Dávila-Acedo, M. A., & Mellado, V. (2017). Teaching technology: From knowing to feeling enhancing emotional and content acquisition performance through Gardner's multiple intelligences theory in technology and design lessons. *Journal of Technology and Science Education, 7*, 58–79.

Santrock, J. W. (2007). *Psikologi Pendidikan*. Jakarta: Kencana Prenada Media Group.

Sheahan, L., While, A., & Bloomfield, J. (2015). An exploratory trial exploring the use of a multiple intelligences teaching approach (MITA) for teaching clinical skills to first year undergraduate nursing students. *Nurse Education Today, 35*, 1148–1154.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False–positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. https://doi.org/10.1177/0956797611417632.

Soleimani, H., Moinnzadeh, A., Kassaian, Z., & Z. & Ketabi, S.. (2012). The effect of instruction based on multiple intelligences theory on the attitude and learning of general English. *English Language Teaching, 5*, 45–53.

Stanciu, D., Orban, I., & Bocos, M. (2011). Applying the multiple intelligences theory into pedagogical practice. Lessons from the Romanian primary education system. *Procedia - Social and Behavioral Sciences, 11*, 92–96. https://doi.org/10.1016/j.sbspro.2011.01.040.

Stanley, T. D., & Doucouliagos, H. (2013). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*, 60–78. https://doi.org/10.1002/jrsm.1095.

Sterne, J. A., Sutton, A. J., Ioannidis, J., Terrin, N., Jones, D. R., Lau, J., et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, 302–307. https://doi.org/10.1136/bmj.d4002.

Teele, S. (2000). *Rainbows of intelligence: Exploring how students learn*. California: Corwin Pres Inc.

Ucak, E., Bag, H., & Usak, M. (2006). Enhancing learning through multiple intelligences in elementary science education. *Journal of Baltic Science Education, 2*, 61–69.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60*, 419–435.

Viechtbauer, W. (2010). Conducting meta–analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03.

Visser, B. A., Asthon, M. C., & Vernon, P. A. (2006a). G and measurement of multiple intelligences: A response to Gardner. *Intelligence, 34*, 507–510. https://doi.org/10.1016/j.intell.2006.04.006.

Visser, B. A., Asthon, M. C., & Vernon, P. A. (2006b). Beyong p: Putting multiple intelligences theory to the test. *Intelligence, 34*, 487–502. https://doi.org/10.1016/j.intell.2006.02.004.

Warren, M. (2018). First analysis of 'pre–registered' studies shows sharp rise in null findings. Retrieved from https://www.nature.com/articles/d41586-018-07118-1.

Waterhouse, L. (2006). Multiple intelligences, the Mozart effect and emotional intelligence: A critical review. *Educational Psychologist, 41*, 207–225. https://doi.org/10.1207/s15326985ep4104_1.

White. (2004). *Howard Gardner: The myth of multiple intelligences*. Lecture presented at the meeting of the Institute of Education University of London. Retrieved from: http://www.ioe.ac.uk/schools/mst/ltv/phil/howardgardnernew_ 171104.pdf.

Willingham, D. T. (2004). *Why don't students like school?* San Francisco: Jossey–Bass.

Xie, J.C. & Lin, R. (2009). Research on multiple intelligences teaching and assessment. Asian Journal of Management and Humanity Sciences, 4, 2-3, 106-24.

Yalmanci, S. G., & Gozum, A. I. C. (2013). The effects of multiple intelligence theory based teaching on students' achievement and retention of knowledge, example of the enzymes subject. *International Journal on New Trends in Education and their Implications, 4*, 2.

Young, B. E. (2003). *Multiple intelligences learning and equity in middle school mathematics education*. Perth, Australia: Doctoral dissertation, Curtin University of Technology. Retrived from https://espace.curtin.edu.au/handle/20.500.11937/2610.