University of Ioannina

School of Education

Department of Primary Education

Natural Sciences in Education

Robotics Interventions for improving educational outcomes - A meta-analysis

Athanasiou T. Lito

Supervisor: Mavridis Dimitris

Ioannina 2017

ABSTRACT

The present study reviews the literature in the field of educational robotics in order to identify potential effects in students' academic performance. After a systematic search in various databases, we included a total of 12 studies in the meta-analysis. The majority of studies were observational and only 4 of them were quasi-experimental. For each study we extract data for the type of robot used, the research method used, the sample characteristics (sample size, age range of students and level of education) and the results observed in terms of scores. We synthesized results from these 12 studies and found that robotic interventions have an overall positive effect in students' academic performance (the overall mean effect was 0.7 and the 95% CI 0.283, to 1.112). All the included studies were assessed with the Cochrane Collaboration 'Risk of bias' tool. Most studies are at high risk of bias and this undermines the validity of results. Large, well-conducted randomized clinical trials are needed to answer the review question conclusively.

1. BACKGROUND

Nowadays we find technology in every aspect of our lives and in various fields such as medicine, industry, home appliances, education and many more. Interest in educational robotics has increased in recent years since students are digital natives and use technology easily. Teachers are trying to include robotics activities in teaching process and many schools and universities offer robotics lessons (in the curriculum) or robotics summer camps due to the benefits they appear to have. It seems that with robots difficult science, engineering and technology abstract concepts can be explained and transformed into real-world understanding. Robots paired with specific software and curriculum offer interesting new learning opportunities and, although they are quite expensive for many school budgets, recent improvements in cost and simplicity make it possible for all students to engage in this kind of hands-on activities.

1.1 How we use robotics

Robotics differs from other modes of learning. It is considered to be an approach that matches constructivism and especially constructionist approaches. Constructionism is a learning and instructional theory which emphasizes the learner's active role and technology is used to create a context that enables "learning by making" and "learning by design". Harel & Papert (1991) suggest that knowledge should not just be transmitted from teacher to pupil, but rather constructed in the mind of the students in the form of active learning. Hence, students should learn with technology and be actively involved in the learning process rather than learn from technology with traditional practices.

As far as it concerns the use of robotics in educational settings, there are different approaches. It can be used as a cognitive tool in a particular lesson, as a teaching technique, or as a subject area itself. Mubin et al. (2013) pointed out the various roles of robot in education – as tutor, tool, or peer. In that way educational robots are categorized as: learning materials, learning companions and teaching assistants.

Some studies use robotics as a cognitive tool. Papert (1993) believed that the computer was a tool that could allow children to explore mathematics and other curricular subjects. Bellou & Mikropoulos (2013) suggest educational robotics as mind tools in Physics and Computer Science education through meaningful learning activities. Educators have started to generate ideas and develop activities to incorporate robotics into the teaching of diverse subjects. STEM (Science, Technology, Engineering and Mathematics) are the subjects that use robotics activities more frequently since concepts in Physics and Technology are relevant to the benefits of educational robotics (Barak & Zadok, 2009). Papert (1980) used Logo programming to teach geometry concepts with the movements of a turtle on the computer screen. That approach has evolved to the visual, drag-and-drop programming languages, such as scratch (MIT) and BYOB (Berkeley University).

Other studies use robotics as a tool to

- teach actual programming languages (Barnes, 2002; Fagin & Merkle, 2003)
- design computer games (Kafai, 1996)
- learn with programmable bricks (Sargent et al., 1996), which led to using products like LEGO Mindstorms for Schools kits (a classic example of robots as learning materials).

1.2 How the intervention might work

Educational theorists such as Papert (1993), believe that robotics activities have tremendous potential to improve classroom teaching. According to Toh et al. (2016) students interacting with robots seem to develop several skills that can be grouped into four major categories: cognitive, conceptual, language and social (collaborative).

The use of robotics in education get students involved in interactive and learning experiences through hands-on experimentation and help children abstract science concepts into real knowledge. Through experimentation, children learn scientific and mathematic principles (Rogers & Portsmore, 2004). According to Nuget et al.(2009) educational robotics supports overall STEM (science, technology, engineering, mathematics) concepts. It also gives students the potential to develop various academic skills and to improve their achievement scores (Barker & Ansorge, 2007; Williams et al., 2007; Highfield, 2010). Another study (Kazakoff et al., 2013) showed positive results in sequencing skills in early childhood children who interact with robots and computer programming. Sullivan (2008) also claims that students who take part in robotics courses and get engaged with robotic activities such as programming and debugging improve their systems understanding and their science process skills. These skills include control of variables, hypothesis generation, hypothesis testing, calculation, construction and evaluation of solutions. Also students in this context need to investigate questions and develop scientific argumentation skills (Baumgartner & Reiser, 1998; Kolodner et al., 2003).

Barack & Zadok (2009) suggest that childrens' involvement with robotics make them come up with more inventive solutions to a problem. In addition, plenty of researches indicate significant positive results in students' problem-solving approaches (Barnes, 1999, Mauch, 2001; Nourbakhsh et al., 2005; Robinson, 2005; Rogers & Portsmore, 2004; Nuget, 2009; Anagnostakis & Michaelides, 2006,). According to Mikropoulos & Bellou (2013) students use robotics as a mind tool to overcome their declarative knowledge and develop procedural knowledge aiming to the solution of a problem.

There are plenty of evidence that robotics indicate equally positive results in children's social skills. It helps them develop teamwork skills (Johnson, 2003) and promotes cooperative learning (Beer et al., 1999; Nourbakhsh et al., 2005). Children develop the spirit of team work and mutual collaboration (Mitnik et al, 2008, Nuget et al, 2009, Owens et al, 2008).

There is no doubt that robotics is the most effective way of motivating students and supporting many subject areas of the curriculum. Robotics make learning fun and sharpen student's thinking. Students see the robots as toys and this makes learning entertaining (Mauch, 2001). Children also increase their imagination due to the fact that these kind of activities require observation, calculation, designing, measurements and testing projects in real life context (Sullivan 2008). According to Johnson (2003) educational robotics encourages children to use their imagination and be innovative.

Finally students develop their language skills in order to achieve a richer interaction (Sugimoto, 2011; Chambers et al., 2008; Bers, 2010; Chang et al., 2010; Young et al., 2010).

1.3 Why it is important to do the review

There are several benefits associated with the use of robotics in a curriculum and there is a rapid development of using technology and multi-media tools in education. Robotics is a trend and they become increasingly popular for educational purposes. Teachers all over the world develop activities and share their ideas to embody robotics into the teaching progress. However, without research evidence to support the direct impact on students' academic performance, robotics activities may be just a trend and could be kept out of classrooms.

Theoretical studies report robotics' benefits in academic performance. However, despite the increasing use of robotics, there is a clear lack of quantitative research on how robotics can increase STEM achievement in students and only a few studies focus on the investigation of the impact in an empirical way. Benitti (2012) points that quantitative analysis is needed. The majority of the literature on educational robotics focuses on describing the activities in robotics educational programs. Only a few studies provide qualitative or quantitative methods to explore the impact of robotics activities and use quantitative measures to evaluate the impact of robotics on student learning.

Hence, are robotics enhancing learning? In order to answer this question and understand the association between using robotics and increased academic performance we performed this systematic review.

2. OBJECTIVES

The purpose of this review is to examine and synthesize empirical evidence of the effectiveness of educational robotics interventions on students' academic performance. Specifically, the primary aim of this review is to answer the following research question: Can educational robotics improve educational outcomes?

Moreover, in this review we aim to assess the effectiveness of using robotic interventions as an educational tool in improving educational outcomes in K-12 students.

3. METHODS

Meta-analysis is a statistical analysis method that combines a collection of analytic results of different studies on a related topic in order to provide effect sizes that can lead to generalizable conclusions about the effect of a treatment (DerSimonian & Laird, 1986; Glass, 1976). These kind of analyses are very commonly used in medical research and lately social sciences adopt them in order to enlighten several fields and to strengthen the evidence that single studies provide. If we test the studies one by one they may be considered too small or to limited, but by combining their findings the results about the treatments efficacy gain more power.

3.1 Criteria for considering studies for this review

3.1.1 Types of studies

Only studies that presented a quantitative assessment of the benefits of robotics in learning were considered. We included both randomized controlled trials (RCT) that used either an experimental or a quasi-experimental (QED) research design and observational studies. We measured the effectiveness of robotics by comparing students who received a robotics program (the experimental group) with a comparison group of students who did not receive any (the control group) or received some kind of comparable control condition. Because of low internal validity, we excluded uncontrolled studies that only reported before and after scores or measurements of academic performance in experimental classes. Finally, we included published and unpublished reports of studies conducted in developed countries after 2000. We rejected studies published before 2000 because we wanted the literature to be relevant to current practices.

3.1.2 Types of participants

We included children in pre-school, primary and secondary school (K-12). We excluded studies conducted in university education. We focused on regular education and alternative education settings from any country.

3.1.3 Types of interventions

We included studies that use robotics as an educational tool for a certain subject. In other words, the objective is not to consider robotics as a teaching subject (such as in robotic courses) but to use robotics as an educational mean to teach another subject (i.e. mathematics, physic, engineering, technology etc.). We included studies that focus on robot or robotics influence on learning, academic performance, pedagogical and developmental domains. We did not used duration criteria, so we included long term and short term studies as long as they were conducted in a school setting (during the school day or in a school-based after school program).

3.1.4 Types of outcome measures

Studies included must report at least one of the following primary outcomes: Achievement scores, academic performance in science concepts and sequencing skills (e.g., standardized achievement tests, mean scores, grades) in major courses (mathematics, physics, language, etc). Also problem-solving abilities in form of scores.

Some secondary outcomes we were interested in are team skills, collaboration, withdrawal and students satisfaction. These outcomes are not included in the meta-analysis but only in the review to help us take an overall picture.

Measurement of the primary outcomes should have been conducted using standardized instruments. We excluded studies that did not provide an effect size, sufficient information or enough data to allow us calculate an effect size.

3.2 Search methods for identification of studies

3.2.1 Electronic searches

We attempted to identify and retrieve both published and unpublished studies that met the inclusion criteria outlined above. The search included multiple electronic databases, research registers, grey literature sources, and reference lists of reviews and relevant studies.

Searches were based on the following search string:

((teaching OR learning OR teach OR learn OR education OR educational) AND (robotic OR robotics OR robot OR robots OR Lego) AND (school OR k-12)).

The search string was adapted for other databases using appropriate controlled vocabulary and syntax.

3.2.2 Electronic databases

The following databases were searched:

(a) IEEE XPLORE,

(b) ACM Digital Library,

(c) ScienceDirect,

(d) ERIC (Educational Resources Information Center)

(e) Scopus

3.2.3 Searching other resources

Grey literature

We searched relative journals to identify relevant unpublished studies and ongoing trials and Conference Proceedings Citation Index to identify conference proceedings. We searched the ProQuest Dissertations & Theses Database to identify relevant dissertations and other unpublished literature.

Reference lists

We searched the reference lists of relevant review articles and included and excluded studies to identify additional studies in the published or unpublished literature. We also conducted forward citation searching using Google Scholar to search for studies citing our included studies.

3.3 Data collection and analysis

3.3.1 Selection of studies

Two review authors screened independently titles and abstracts. We excluded studies with a title or an abstract that was irrelevant to the review question and did not satisfy the inclusion criteria. We resolved any disagreement by looking at the relevant manuscripts and by discussion. For eligible studies, full papers were retrieved and judged independently by the authors to identify those satisfying the inclusion criteria. In case of uncertainty or discrepancy, concurrence resolved through discussion or by consulting a third assessor in order to reach a consensus about the study's eligibility.

3.3.2 Data extraction and management

For eligible studies that met the inclusion criteria and passed the screening stage, two review authors extracted the data using data extraction forms.

We extracted data concerning:

General information (title, published/unpublished, authors, year of publication, country, date of data extraction, sponsors)

Participants (sample size/number or participants randomized to the study, number of withdrawals, age/grade, gender and ethnicity)

Intervention (type(s), robot types used in the research, duration and intensity of intervention)

Outcome (outcome measures, the subject that the researcher wanted to teach through robotics, primary/secondary outcomes, effect sizes)

We resolved discrepancies through discussion or, when required, we consulted a third person. We present the data in "4.3 Characteristics of studies".

3.3.3 Assessment of risk of bias in included studies

Two review authors coded each included study using the "Cochrane tool for assessing risk of bias" (Higgins & Green, 2008). This includes the assessment of selection bias (random sequence allocation and allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment and incomplete outcome data), reporting bias (selective outcome reporting) and other sources of bias. We reported any study characteristic that seems peculiar and may affect the magnitude of effect for the robotic interventions. The review authors judged the risk of bias as either 'high risk' of bias, 'low risk' of bias or 'unclear risk' of bias and each study was coded as "low", "high", or "unclear" risk of bias on each of the domains. In case of discrepancies the two reviewers resolved them through consensus. If consensus could not be reached between the two reviewers, a third member of the review team was consulted.

We also coded for additional factors related to potential bias in this corpus of studies: researcher allegiance bias and funding source bias. We assessed whether the study authors were directly involved or were funded by an entity that had some stake in the intervention. We believe it was important to assess for these biases because studies are more likely to be biased in favor of the treatment intervention when study authors have a direct role in the study or when studies are funded by a source that takes part in the intervention (Lundh et al., 2012).

We provide details on each study's risk of bias in a 'Risk of bias' table, with information provided that led to each judgement.

3.3.4 Synthesis procedures and statistical analysis

In this review we conducted descriptive analyses on variables to provide information. The variables of our interest regard:

- participants (gender, grade, age, subgroups)
- settings where studies took place (school type, classroom type, country, during the school day or in a school-based after school programs)
- risk of bias across studies included in the meta-analyses

We calculated overall effects from the trials where data are available.

Following descriptive analysis, we conducted quantitative analysis. Evidence were synthesized in a quantitative way (meta-analysis), where possible. We calculated effect sizes for each included study when data were reported in the study or provided by study authors. The studies involved in the meta-analysis had continuous outcomes thus we analyzed them using the Standardized Mean Difference (SMD) with 95% confidence intervals (CI). Mean differences and 95% CI were calculated comparing and pooling the mean score differences from the end of treatment to baseline for each group.

Some investigators present different instruments to measure outcomes, either because they use different definitions of a particular outcome or because they choose different instruments to measure the same outcome. We used the Standardized Mean Differences (SMD) and 95% confidence intervals to combine trials that assess the same outcome with different measures or instruments. In cases of combining two reported subgroups into a single group, we followed Cochrane Handbook's (2011) recommendations about combining groups (cite, 7.7.3.8)

When there was evidence of skewed data, it was reported. In case of missing standard deviations of the change score we used final values and the respective standard deviations for each group. We assumed that data not reported for some outcomes or groups are not missing at random and that missing cases had poor outcomes.

For our meta-analysis we used three methods to identify statistical heterogeneity:

- ✓ creating forest plots to see if the confidence intervals of individual studies have poor overlap. We constructed forest plots displaying study-level mean effect sizes and 95% confidence intervals for the included studies to provide opportunity for visual analysis of the precision of the estimated effect sizes, detection of studies with extreme effects, and information regarding heterogeneity of studies.
- ✓ conducting a χ^2 test to compare the observed variance to what would be expected from sampling error. We considered a meta-analysis to have heterogeneity if its χ^2 P value is less than 0.10.
- ✓ calculating an I^2 statistic to describe the percentage of total variation across studies due to the heterogeneity rather than chance. We considered a meta-analysis to have heterogeneity if I^2statistic is greater than 50%.

In order to discover the cause of heterogeneity we conducted subgroup and sensitivity analyses.

We conducted the following subgroup analyses and a χ^2 test for each analysis to determine whether or not the effects of robotics are statistically significantly different for different subgroups:

- Age- grade (elementary school, pre K-K, middle school).
- Intervention duration (less than 1 month, 1 to 6 months, more than 6 months).

In sensitivity analysis we investigated the influence of study characteristics on the robustness of the review results. More specifically, we conducted the following sensitivity analyses:

- studies at high/unclear risk of bias for allocation concealment and sequence generation and studies at low risk of bias.

To synthesize effects across studies, a weighted mean effect was calculated by weighting each study level effect size by the inverse of its variance.

Small-study effects was assessed using funnel plots and the Egger's test and was considered a proxy for publication bias.

3.4 Deviations from the protocol

There were certain points in this review that required deviation from the protocol. During the electronic databases search the SpringerLink, the Wilson Education and the MIT were not accessible thus we did not search those databases as planned.

The majority of the included studies did not report any of the secondary outcomes (team skills, collaboration, withdrawal and students satisfaction) so we were not able to draw conclusions about these outcomes.

We did not executed a sub-group analysis for different genders due to lack of clear data in some studies.

4. RESULTS

4.1 Results of search

Our search of databases and other sources (journal and references) identified a total of 1927 citations. The first step was to screen titles and abstracts. 1896 studies were excluded due to lack of relevance. The remaining 31 studies were retrieved and their full text was screened for eligibility. Twelve studies passed full-text screening and were included in the review. Figure 1 demonstrates a flow chart of the study selection process. Also Table 1 shows the study selection from the databases.

Figure 1: flow chart of study selection process

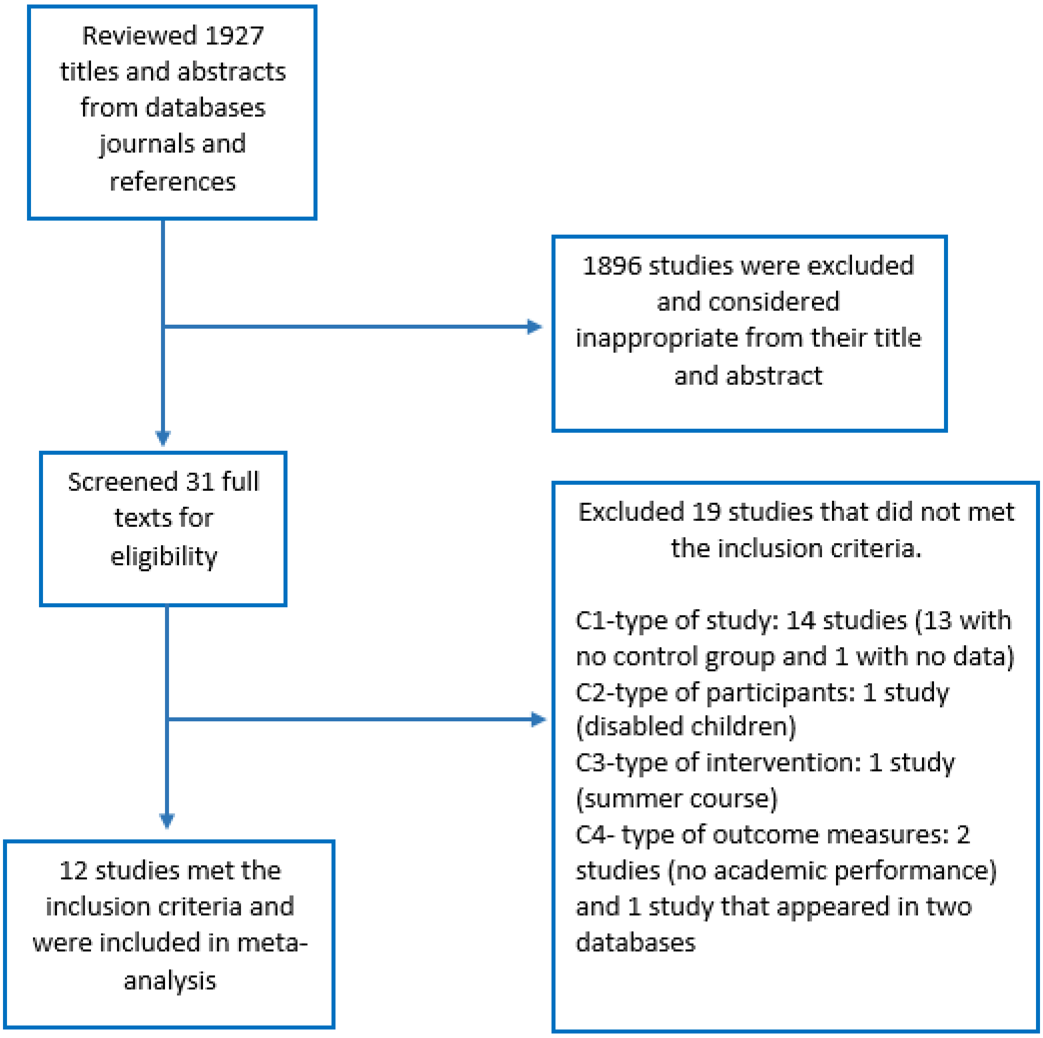


Table 1: Study selection from each database.

Database	Search results	Relevant title and article	Included in meta-analysis
IEEE	3085	3	1
ACM	475	4	0
ERIC	563	22	7
Scopus	498	2	2
ScienceDirect	83	2	2

4.1.1 Included studies

All of the 12 studies that met criteria for inclusion in this review were QED studies and provided enough information to compute an effect size. All of them are included in our meta-analyses. Analytical characteristics of the included studies are reported.

4.1.2 Excluded studies

Nineteen studies were excluded during the stage of full text screening. Most of the studies were excluded due to study design characteristics (13 studies had no control group for comparison and 1 study had no sufficient data). The rest of them were excluded due to not meeting criteria related to intervention characteristics (1 study with summer course), type of participants (1 study with disabled children) and type of outcome measures (2 studies that did not measure academic performance). Finally 1 study was duplicate in two different databases.

4.2 Risk of bias in included studies

- Random sequence generation:

  The majority of the studies (n=7, 58.3%) did not provide information about the method of randomization used in the study. We assessed these studies as unclear. Only one study (8.3%) was assessed to be at low risk of bias due to description of the randomization. We rated the remaining 4 studies (33.3%) as high risk of bias due to the non-random selection of participants.

- Allocation concealment:

  We rated all of the studies (100%) as high risk of bias due to luck of allocation concealment.

- Blinding of participants & those delivering intervention:

  In both of the categories we rated all the studies (100%) as high risk of bias. The study design and the characteristics of these interventions cannot support blinding of participants and teachers.

- Blinding of outcome assessors:

  Although none of the studies clearly indicates that outcome assessors were blinded we deemed that measurements were unlikely to be influenced. In this case we rated all the studies (100%) as low risk of bias.

- Incomplete outcome data:

  The majority of the studies (n=7, 58.3%) reported no missing data and were rated as low risk of bias. The researchers were unable to collect all the data due to dropouts from four studies (33.3%) but the number of dropouts is not mentioned. We rated these studies as unclear. One study (8.3%) provided details about incomplete outcome data but was rated as high risk due to the big percentage of dropouts.

- Selective reporting:

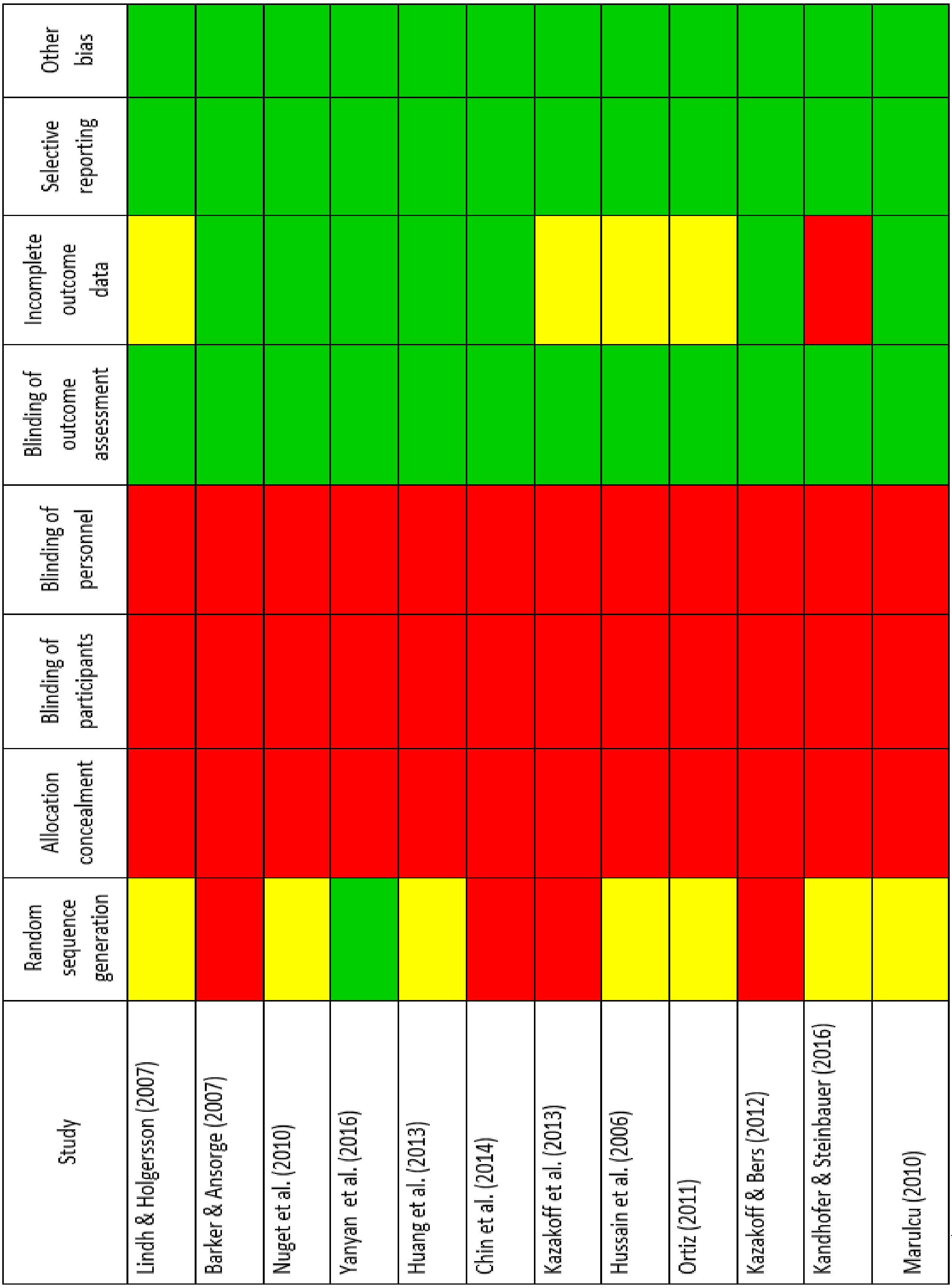
  The majority of the studies (n=11, 91.3%) report all the pre specified outcome measures and were rated low risk of this bias. We rated only one study (8.3%) as unclear since all the expected outcomes were not reported clearly at the beginning.

- Other risks of bias:

  There were no other risks of bias in all of the studies.

See Table 2 for a summary of risk across studies.

Table 2: 'Risk of bias' summary- review authors' judgements about each risk of bias item for each included study.



4.3 Characteristics of studies

Lindh & Holgersson (2007)

Country	Sweden	
Methods	QED	
Participants	Age/grade: fifth grade (12-13 years old) and ninth grade (15-16 years old) N: 696 Gender: --	
Interventions	Intervention description/ robot used: LEGO kit Duration of interventions: 2 hours per week during 12 months Control condition: ordinary teaching	
Outcomes	Mathematics tests Problem solving tests Attitudes	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated. Unclear effect on bias.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were aware of treatment group.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias)	Unclear risk	Information not available
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Barker & Ansorge (2007)

Country	United States, Nebraska	
Methods	Observational study	
Participants	Age/grade: mean age was 9.00 N: 32 Gender: 64% male, 36% female	
Interventions	Intervention description/ robot used: LEGO kit, ROBOLAB software Duration of interventions: one hour twice a week for six weeks Control condition: traditional teaching	
Outcomes	Achievement in science, engineering and technology	
Risk of bias		
Bias	**Authors' judgement**	Support for judgement
Random sequence generation (selection bias)	High risk	Participants of the experimental group were the participants of the afterschool program. Only the comparison group was randomly selected from the remaining students of the class.
Allocation concealment (selection bias)	High risk	Non-random, predictable sequence.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were aware of treatment group. Unclear effect on bias.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Nuget et al. (2010)

Country	United States, Nebraska	
Methods	QED	
Participants	Age/grade: mean age was 12.28 N: 288 Gender: 76% male, 24% female	
Interventions	Intervention description/ robot used: LEGO Mindstorms NXT robotics platform Duration of interventions: 40 hours Control condition: (a) control group with no intervention (b) sort term intervention – 3 hours	
Outcomes	STEM learning and attitudes	
Risk of bias		
Bias	**Authors' judgement**	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were aware of treatment group.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Yanyan et al. (2016)

Country	China, Beijing	
Methods	QED	
Participants	Age/grade: fourth grade (10 years old) N: 30 Gender: 20 male, 10 female	
Interventions	Intervention description/ robot used: LEGO bricks Duration of interventions: 2 hours per week for five weeks Control condition: commonly used pedagogy	
Outcomes	Science performance and problem solving abilities	
Risk of bias		
Bias	**Authors' judgement**	Support for judgement
Random sequence generation (selection bias)	Low risk	Students were randomly divided into two groups.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Huang et al. (2013)

Country	______	
Methods	QED	
Participants	Age/grade: fifth and sixth grade N: 80 Gender: --	
Interventions	Intervention description/ robot used: LEGO Duration of interventions: 7 weeks Control condition: flow-charts	
Outcomes	Programming skills Attitude Interaction	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	Unclear risk	The way that the researchers divided the groups in not mentioned.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Unclear risk	Not all the expected outcomes were reported clearly at the beginning.
Other bias	Low risk	The study appears to be free of other sources of risk

Chin et al. (2014)

Country	Taiwan	
Methods	Observational study	
Participants	Age/grade: mean age 8.7 years N: 52 Gender: 27 male, 25 female	
Interventions	Intervention description/ robot used: Educational robot-based learning system Duration of interventions: 6 weeks Control condition: PowerPoint-based learning system	
Outcomes	Learning performance and students’ motivation.	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	High risk	Non-random selection of participants. One class was experimental group and the other the control group.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Kazakoff et al. (2013)

Country	United States	
Methods	Observational study	
Participants	Age/grade: pre-kindergarten and kindergarten N: 42 Gender: --	
Interventions	Intervention description/ robot used: CHERP software and LEGO WeDo Duration of interventions: 1 week Control condition: no robotics interaction	
Outcomes	Sequencing ability	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	High risk	Non-random selection of participants.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	The researchers were unable to collect all the data due to dropouts. The number of dropouts is not mentioned. Unclear effect on bias.
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Hussain et al. (2006)

Country	Sweden	
Methods	QED	
Participants	Age/grade: fifth and ninth grade (12-13, 15-16 years old) N: 696 Gender: --	
Interventions	Intervention description/ robot used: LEGO kit Duration of interventions: 2 hours a week for 12 months Control condition: traditional learning activities	
Outcomes	Achievement in mathematics	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias)	Unclear risk	Information not available
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Ortiz (2011)

Country	_____	
Methods	QED	
Participants	Age/grade: second grade N: 30 Gender: --	
Interventions	Intervention description/ robot used: LEGO kit Duration of interventions: 15 hours Control condition: school text book	
Outcomes	Achievement in computation, engineering and mathematic.	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias)	Unclear risk	Information not available
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Kazakoff & Bers (2012)

Country	United States, Boston	
Methods	Observational study	
Participants	Age/grade: kindergarten N: 54 Gender: --	
Interventions	Intervention description/ robot used: CHERP Duration of interventions: twice a week (60-90 minutes at a time) Control condition: typical curriculum	
Outcomes	Sequencing skills	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	High risk	Mixed method of randomization in several classes (some of them are random but others not).
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias)	Low risk	54/58 students provided post intervention data due to excessive absences from activities. Reasons for missing data not related to outcome.
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Kandhofer & Steinbauer (2016)

Country	Austria and Sweden	
Methods	QED	
Participants	Age/grade: mean age 14.9 N: 148 Gender: 40% female, 57% male and 3% not stated	
Interventions	Intervention description/ robot used: LEGO Mindstorms NXT platform Duration of interventions: weekly robotic courses (approx. 8 months) Control condition: science courses	
Outcomes	Technical skills Attitudes Interest in science	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias)	High risk	148/242 students provided post intervention data. Classified as high risk due to the big percentage.
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

Marulcu (2010)

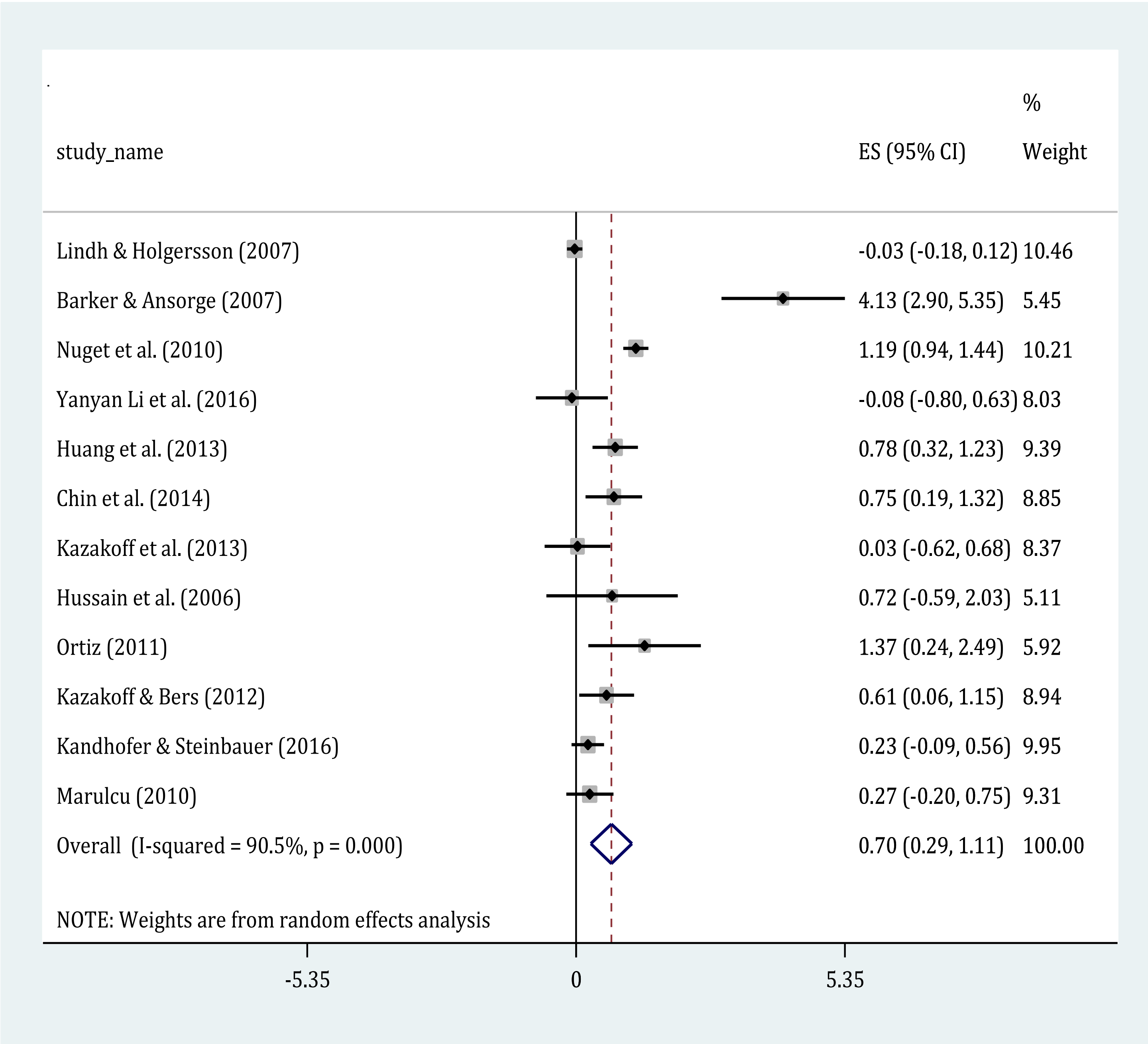
Country	United States, Boston	
Methods	QED	
Participants	Age/grade: fifth grade N: 79 Gender: --	
Interventions	Intervention description/ robot used: LEGO engineering-design practices Duration of interventions: a year Control condition: science courses, Full Option Science System's (FOSS)	
Outcomes	Understanding of science content	
Risk of bias		
Bias	**Authors' judgement**	**Bias**
Random sequence generation (selection bias)	Unclear risk	The method of randomization is not described in the paper.
Allocation concealment (selection bias)	High risk	Random sequence known to staff in advance.
Blinding (performance bias and detection bias) participants	High risk	Participants were aware of the group they were allocated.
Blinding (performance bias and detection bias) those delivering intervention	High risk	Those delivering the interventions were not blinded.
Blinding (performance bias and detection bias) outcome assessors	Low risk	No blinding, but measurement unlikely to be influenced
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Low risk	From the study report, it appears that all the prespecified outcome measures have been reported.
Other bias	Low risk	The study appears to be free of other sources of risk

4.4 Synthesis of results

4.4.1 Mean effects on academic outcomes

To explore the efficacy of robotic interventions on improving educational outcomes we synthesized effect sizes from 12 studies. Results indicate that the overall mean effect was 0.70 and the 95% CI 0.28 to 1.11 demonstrating an overall positive and statistically significant effect of interventions on academic outcomes. We present the mean effects and confidence intervals for each study separately in the forest plot below (Figure 2)

Figure 2: Forest plot of mean effects on academic outcomes

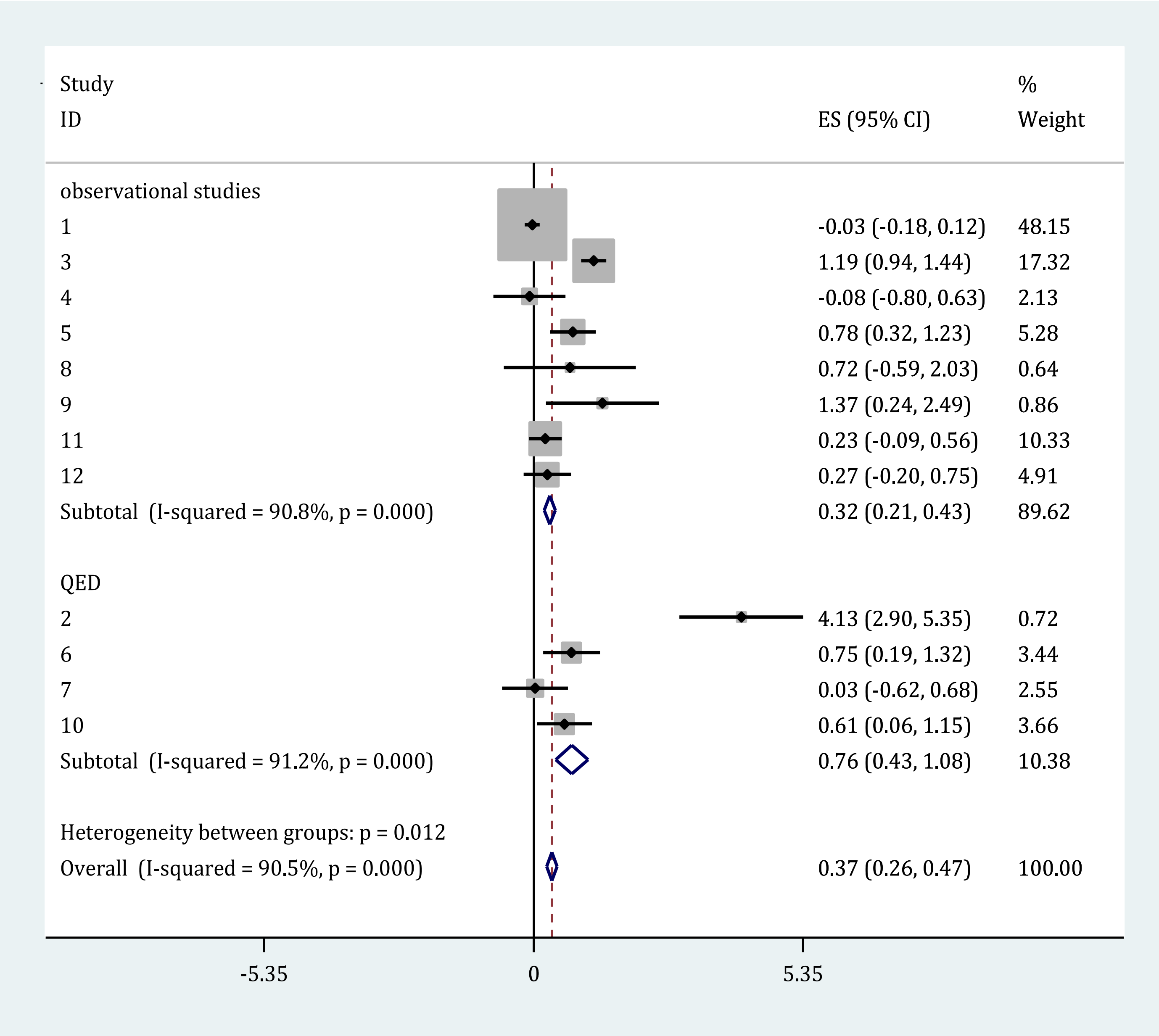


We noticed overall positive and statistically significant effect of robotics interventios on academic performance outcomes. Heterogeneity analysis indicated a considerable amount of heterogeneity that was statistically significant (I^2 =90.5 %, Q = 116.14, p = .00, Tau2 = .4157). Although the grand mean effect size provides evidence that the robotic interventions were, on average, effective, the highly heterogeneous nature of the distribution suggests large differential effects across studies. In order to predict an estimate of an interval in which future observations will fall we also calculated the 95% prediction interval (-0.628 to 2.028). The predictive interval shows the plausible range of values for the effect size in a future study. It includes zero suggesting that future trials may have a zero effect and the overall effect may move towards zero in the future. Hence, although we found a statistically significant positive effect, results are not conclusive due to the large amount of heterogeneity.

4.4.2 Sensitivity analysis

We conducted sensitivity analyses to investigate the influence of study characteristics on the robustness of the review results. We comment on how these sensitivity analyses influence pooled effect size. For the 4 studies with inadequate allocation concealment and therefore at high risk of bias, the SMD was 0.76 (95% CI 0.44 to 1.08). For the 8 studies with unclear of low risk of allocation concealment the SMD was 0.32 (95% CI 0.21 to 0.43) (Figure 3). Comparing the two size effects we note that the diamonds don't overlap, but they only have one common point at 0.43. That indicates significant difference between studies at high versus low/unclear risk of bias judgements. Surprisingly, it is the studies at low risk of bias for allocation concealment that have a larger effect.

Figure 3: Sensitivity analysis



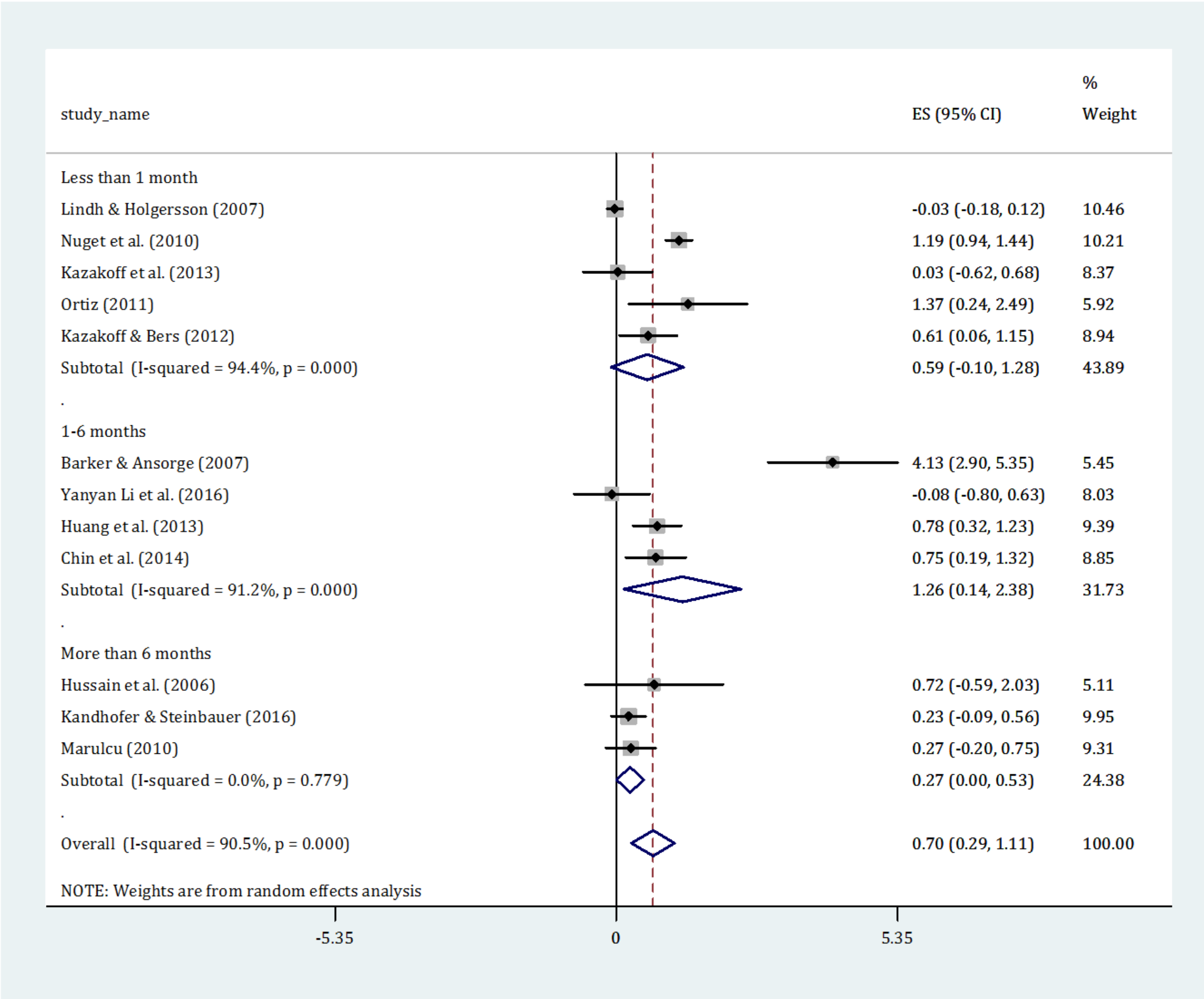
4.4.3 Subgroup analysis

To investigate heterogeneity furthermore we split participants into subgroups to make comparisons between them.

Duration of intervention

We explored the influence of the total duration of the interventions (less than a month, 1-6 months, 6 or more months) on the improvement of academic performance for the intervention groups comparing to the control groups (Figure 4). A large effect size was observed for less than 1 month interventions (SMD 0.59, 95% CI -0.10 to 1.28). A large effect size was observed also for 1-6 months interventions (SMD 1.26, 95% CI 0.14 to 2.38) and finally the effect size of more than 6 months interventions was SMD 0.27, 95% CI 0.00 to -0.53. From the figure below a large overlapping is obvious and the heterogeneity of the subgroups ranges at high percentages.

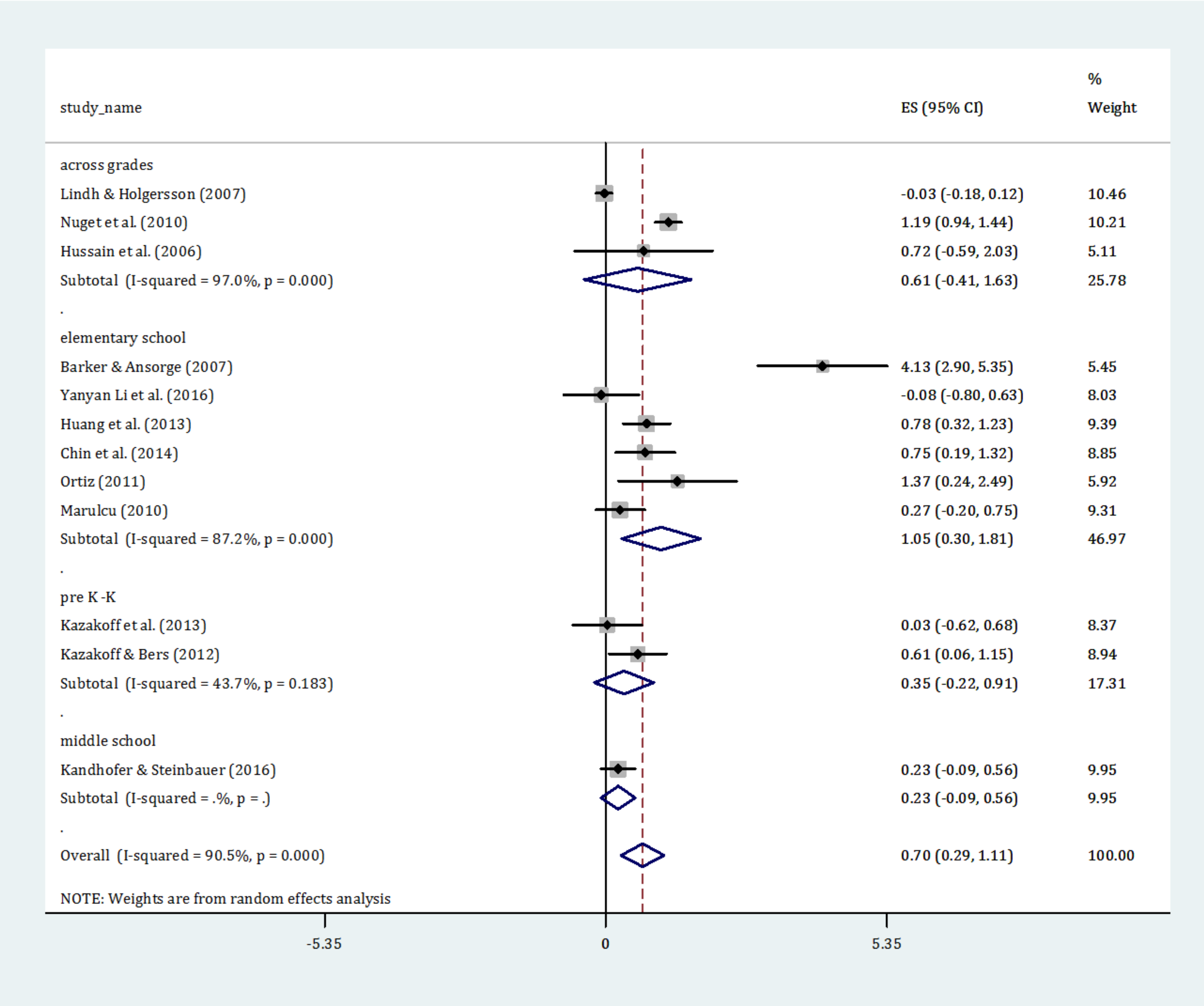
Figure 4: Subgroup analysis - duration of the intervention



Grade of participants

We explored the influence of the grade of participants (preK –K, elementary school, middle school, across grades students) on outcomes referring to academic performance for intervention groups versus control groups (Figure 5). The effect size of across grades students was 0.61, 95% CI -0.417 to 1.63, of elementary school students 1.05, 95% CI 0.30 to 1.81, of pre K –K students 0.35 95% CI -0.22 to 0.91 and of middle school students 0.23 95% CI -0.09 to 0.56. We notice no significant difference at the subgroup analysis' results and the heterogeneity remains high, with only exception the sub-group preK-K.

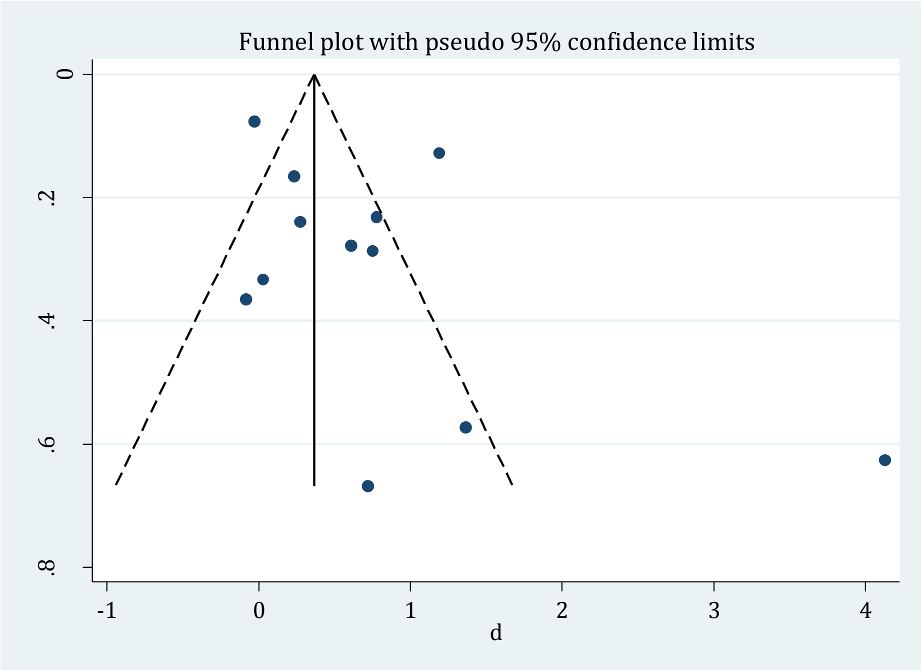
Figure 5: Subgroup analysis – grade of participants



4.5 Small-study effects

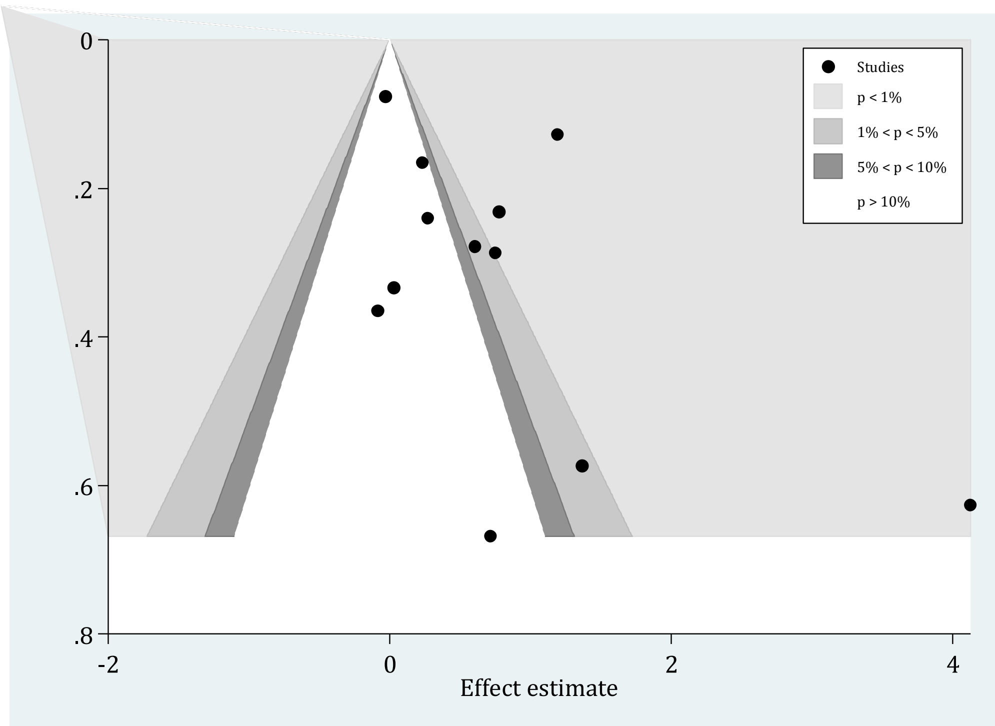
We explored for small-study effects by drawing a funnel plot. Small-study effect are typically used as a proxy for publication bias.

Figure 6: Funnel plot of comparison robotics vs. control group outcome



The funnel plot, as shown in Figure 6, is asymmetric and a few smaller studies tended to exaggerate the effectiveness of intervention, indicating the possibility of publication bias. The contour-enhanced funnel plot (Figure 7) helped us distinguish between publication bias and other causes of the asymmetry such as heterogeneity. It showed that studies were distributed not in both statistical significance and in non-statistical significance areas (grey and white).

Figure 7: Contour-enhanced funnel plot



We also run an Egger's meta-regression model (Table 3). The CI of bias includes the zero value and p=0.121 so we cannot conclude that a significant difference exists.

We finally applied the trim-and-fill method (Table 4). The addition of the 6 estimated unpublished studies moved the effect size to 0.104 with 95% CI -0.332 to 0.540 (p=0.640).

Table 3: Egger's test

Number of studies = 12   Root MSE = 3.004

Std_Eff	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
slope	-.0159666	.2763157	-0.06	0.955	-.6316363	.5997031
bias	2.534434	1.49658	1.69	0.121	-.8001533	5.869021

Test of H0: no small-study effects P = 0.121

Table 4: Trim-and-fill method

Filled
Meta-analysis

Method	Pooled Est	95% CI Lower	95% CI Upper	Asymptotic z_value	Asymptotic p_value	No. of studies
Fixed	0.064	-0.028	0.156	1.355	0.175	18
Random	0.104	-0.332	0.540	0.468	0.640	

Test for heterogeneity: Q= 289.114 on 17 degrees of freedom (p= 0.000)

We decided to exclude Barker & Asgore's study due to outliers and re-run Egger's test and trim and fill method. The p value increases (p=308) and we still cannot conclude that there is evidence of publication bias. The results indicate a slight change of the effect sizes (Tables 5-6) that does not help us to interpret the results.

Table 5: Eggers test without Barker & Asgore's study

Number of studies = 11   Root MSE = 2.803

Std_Eff	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
slope	.0983709	.2678379	0.37	0.722	-.5075206	.7042623
bias	1.630158	1.509681	1.08	0.308	-1.784978	5.045295

Test of H0: no small-study effects P = 0.308

Table 6: Trim and fill method without Barker & Asgore's study

```
Filled
Meta-analysis

       |  Pooled      95% CI        Asymptotic      No. of
Method |     Est   Lower   Upper  z_value  p_value   studies
-------+----------------------------------------------------
Fixed  |   0.064  -0.029   0.157    1.347    0.178     16
Random |   0.109  -0.285   0.502    0.542    0.588

Test for heterogeneity: Q= 205.123 on 15 degrees of freedom (p= 0.000)
Moment-based estimate of between studies variance =  0.530
```

5. DISCUSSION

5.1 Summary of main results

The present review includes a total of 12 studies, 8 observational and 4 quasi-experimental, that met the inclusion criteria and examine the effects of educational robotic intervention on academic performance and problem solving abilities. All of the included studies provided the necessary data that are combined at the meta-analysis. Results of the meta-analysis indicate small positive effects. It seems that educational robotics may benefit students' academic performance but the heterogeneity of the outcomes is high and compromised the validity of results. That can be explained from the high diverse in the studies characteristics and the diverse nature of the studies, both in terms of design and methods employed. The review of the studies indicates a lot of difference among studies as far as intervention's duration, type of robotics used, age- grade applied etc. The sub group analysis that was conducted in order to clarify these differences also shown high heterogeneity in the majority of the sub-groups, which lead to the conclusion that we dealt with poor quality studies. It is really common in social studies that there is not much explanation in methodological details and some fields were unclear, for example a lot of the studies mention the intervention's duration but not the frequency of the total hours or mention about students that are not included in the results but do not explain reasons of drop out. Given the quality and high risk of bias across studies in several areas, caution must be used in the interpretation of the study results.

The sensitivity analysis indicate difference between quasi- experimental studies and observational studies as there was no overlapping at the forest plot. The quasi- experimental studies, as expected, gave larger effect size. It is important to mention the low number or quasi-experimental studies and the total luck or control randomized trials.

The results of the sub-group analysis also show that long term interventions have greater impact and present better outcomes, with largest effect size at the group of 1-6 months interventions. But not very long term while at the group of 6 months and more the effect size decreases. This can be explained because academic outcomes are often more difficult to change immediately, and given that the studies in this review measured grades, it may take a longer periods to see change in grades. On the other hand students in the long term interventions may lose their interest.

The sub-group analysis for the students' grade gave larger effect size at the group of elementary school students. Also as it was expected from the literature the majority of studies are carried out at elementary schools. The other groups also have positive effects which indicate that robotic interventions effect academic performance in all grades (K-12).

5.2 Quality of the evidence

The risk of bias summary gives information about the quality of the evidence. Some important risk of bias present in the majority of the studies, related to allocation concealment and blinding. That was expected due to the nature of these kind of studies. Most of them do not use randomization and blinding is very difficult to be achieved as it is obvious to the stuff and to the students that they use robotics. These two fields (allocation concealment and blinding) are the prime reasons that threatens the internal validity of the included studies and we are concerned that this body of evidence is biased in favor of the robotics due to experimenter expectancy effects. That leads to caution regarding the meta-analysis results' use and interpretation.

In all of the studies selective reporting was rated as low risk of bias. Also some of the included studies were sponsored by organizations or authorities that are not relevant in robotics industry and we are not concerned that this caused bias in favor of robotics.

Finally the funnel plot was also asymmetrical which indicates potential for publication bias.

5.3 Limitations and potential biases in the review process

During the searching for the included studies we faced difficulties, as expected from the literature, such as the low number of relevant studies and the luck of studies with enough empirical data and comparison between treatment and control groups. We made every attempt to search for published studies, however, the majority of the studies were published journal articles and only one of them was an unpublished dissertation. Another limitations was the geographical restriction and some not accessible databases that may gave us further studies and more information. Finally there is some indication of publication bias present, which could be upwardly biasing the mean effect for the outcomes examined in this review.

6. REFERENCES

6.1 References to included studies

Barker, B. S., & Ansorge, J. (2007). *Robotics as means to increase achievement scores in an informal learning environment*. Journal of Research on Technology in Education, 39(3), 229–243.

Chin, K. Y., Hong, Z. W., & Chen, Y. L. (2014). *Impact of Using an Educational Robot-Based Learning System on Students' Motivation in Elementary Education*. IEEE Transactions on Learning Technologies, vol. 7, pp. 333-345.

Huang, K., Yang, T., Cheng, C. (2013). *Engineering to see and move: Teaching computer programming with flowcharts vs Lego robots*. iJET, 8(4).

Hussain, S., Lindh, J., & Shukur, G. (2006). *The effect of LEGO training on pupils' school performance in mathematics, problem solving ability and attitude: Swedish data.* Journal of Educational Technology and Society, 9(3), 182–194.

Kandlhofer, M., & Steinbauer, G. (2016). *Evaluating the impact of educational robotics on pupils' technical- and social-skills and science related attitudes*. Robotics and Autonomous Systems, 75, 679-685.

Kazakoff, E., & Bers, M. (2012). *Programming in a robotics context in the kindergarten classroom: The impact on sequencing skills*. Journal of Educational Multimedia and Hypermedia, 21(4), 371–391.

Kazakoff, E., Sullivan, A., & Bers, M. (2012). *The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood*. Early Childhood Education Journal, 41(4), 245–255.

Li, Y., Huang, Z., Jiang, M., & Chang, T. W. (2016). *The Effect on Pupils' Science Performance and Problem-Solving Ability through Lego: An Engineering Design-based Modeling Approach*. Educational Technology & Society, 19 (3), 143–156.

Lindh, J., & Holgersson, T. (2007). *Does lego training stimulate pupils' ability to solve logical problems?* Computers & Education, 49(4), 1097–1111.

Marulcu, I. (2010). *Investigating the impact of a LEGOTM-based, engineering-oriented curriculum compared to an inquiry-based curriculum of fifth graders' content learning of simple machines*. Dissertation Abstracts International (UMI No. 3419134)

Nugent, G., Barker, B., Grandgenett, N., & Adamchuk, V. I. (2010). *Impact of robotics and geospatial technology interventions on youth STEM learning and attitudes*. Journal of Research on Technology in Education, 42(4), 391–408.

Ortiz, A.M., (2011). *Fifth Grade Students' Understanding of Ratio and Proportion in an Engineering Robotics Program.* Proc. Amer. Soc. Eng. Ed. Session M444B. British Columbia, Canada.

6.2 Additional references

Anagnostakis, S., Michaelides, P.G. (2006). *Laboratory of educational robotics' - an undergraduate course for primary education teacher – students.* In: 3rd International Conference on Handson Science, Braga, Portugal, pp. 329–335

Barak, M., & Zadok, Y. (2009). *Robotics projects and learning concepts in science, technology and problem solving.* International Journal of Technology and Design Education, 19(3), 289–307.

Barnes, D. J. (2002). *Teaching introductory Java through Lego Mindstorms models*. Proceedings of the 33rd SIGCSE technical symposium on computer science education.

Barnes, M. (1999). *Cumulative and Exploratory Talk in a Collaborative Learning Classroom.* Proceedings of the 22nd Conference of the Mathematics Education Research Group of Australasia 53-59, Adelaide, South Australia.

Baumgartner, E. & Reiser, B. (1998). *Strategies for supporting student inquiry in design tasks*. Presented at the Annual Meeting of the American Educational Research Association, San Diego, California.

Beer, R. D., Chiel, H. J., & Drushel, R. F. (1999). *Using robotics to teach science and engineering*. Communications of the ACM, *42*(6), 85-92.

Benitti, F.B.V. (2012). *Exploring the educational potential of robotics in schools: a systematic review*. Computers and Education, 58(3), 978–988.

Bers, U. M., Ponte, I., Juelich, C., Viera, A., Schenker, J. (2002). *Teachers as designers: integrating robotics in early childhood education*. Information Technology in Childhood Education Annual, 14, 123-145.

Chambers, J. M., Carbonaro, M., & Murray, H. (2008). *Developing conceptual understanding of mechanical advantage through the use of Lego robotic technology.* Australasian Journal Educational Technology, 24(4), 384-401.

Chang, C.W., Lee, J.H., Chao, P.Y., Wang, C.Y. (2010). *Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school*. Educational Technology and Society, 13(2), 13–24.

DerSimonian R, Laird N.(1986) *Meta-analysis in clinical trials.* Controlled Clinical Trials 7:177–188

Fagin, B., & Merkle, L. (2003). *Measuring the effectiveness of robots in teaching computer science.* Proceedings of the 34rd SIGCSE technical symposium on computer science education.

Glass, G. V. (1976) *Primary, secondary and meta-analysis*. Educational Research, 5, 3–8.

Higgins, J. P. and Altman, D. G. (2008) Assessing Risk of Bias in Included Studies, in Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series (eds J. P. Higgins and S. Green), John Wiley & Sons, Ltd, Chichester, UK

Highfield, K. (2010). *Robotic toys as a catalyst for mathematical problem solving*. Australian Primary Mathematics Classroom, 15(2), 22-27.

Johnson, J. (2003). *Children, robotics and education*. In Proceedings of 7th international symposium on artificial life and robotics (Vol. 7, pp. 16–21), Oita, Japan.

Kafai, Y. (1996). *Learning Design by Making Games: Children's Development of Design Strategies in the Creation of a Complex Computational Artifact*. Constructionism in practice: Designing, thinking, and learning in a digital world, Mahwah, New Jersey, USA: Lawrence Erlbaum.

Kazakoff, E. R., Sullivan, A., & Bers, M. U. (2013). *The Effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood*. Early Childhood Educational Journal, 41, 245–255.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003) *Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design™ into practice.* Journal of the Learning Sciences 12 (4): 495–547.

Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O. A., & Bero, L. A. (2012). *Industry sponsorship and research outcome*. Cochrane Database of Systematic Reviews.

Mauch, E. (2001). *Using technological innovation to improve the problem solving skills of middle school students.* The Clearing House, 75(4), 211–213.

Mikropoulos, T. A., & Bellou, I. (2013). *Educational Robotics as Mindtools*. Themes in Science and Technology Education, 6(1), 5-14.

Mitnik, R., Nussbaum, M., & Soto, A. (2008). *An autonomous educational mobile robot mediator.* Autonomous Robots, 25(4), 367–382.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). *A Review of the applicability of robots in education*. Technology for Education and Learning, 1, 1-7.

Nuget, G., Barker, B., Grandgenett, N., & Adamchuk, V. (2009). *The Use of digital manipulatives in K-12: robotics, GPS/GIS and programming.* Proceedings of the 39th IEEE international conference on Frontiers in education conference (pp. 302-307) San Antonio Texas: IEEE Press.

Nourbakhsh, I., Crowley, K., Bhave, A., Hamner, E., Hsium, T., Perez-Bergquist, A., Richards, S., & Wilkinson, K. (2005). *The robotic autonomy mobile robots course: Robot design, curriculum design, and educational assessment.* Autonomous Robots, 18(1), 103–127.

Owens, G., Granader, Y., Humphrey, A., & Baron-Cohen, S. (2008). *LEGO Therapy and the Social Use of Language Programme: An Evaluation of Two Social Skills Interventions for Children with High Functioning Autism and Asperger Syndrome.* Journal of Autism and Developmental Disorders, 38(10), 1944–1957.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* New York: Basic Books, Inc.

Papert, S. (1993). *The Children's Machine: Rethinking School in the Age of the Computer.* New York: Basic Books.

Papert, S., & Harel, I. (1991). *Situating constructionism. Constructionism.* New York: Ablex Publishing Corporation: 193-206.

Robinson, M. (2005). *Robotics-driven activities: Can they improve middle school science learning?* Bulletin of Science, Technology & Society, 25(1), 73-84.

Rogers, C., & Portsmore, M., (2004) *Bringing Engineering to Elementary School*. Journal of STEM Education. Vol. 5(3/4), pp.17-28.

Sargent, R., Resnick, M., Martin, F., and Silverman, B. (1996). *Building and Learning with Programmable Bricks.* Constructionism in Practice, Hillsdale, NJ.

Sugimoto, M. (2011). *A Mobile mixed-reality environment for children's storytelling using a handheld projector and a robot*. IEEE Trans Learning Technologies, 4(3), 249-260.

Sullivan, F. R. (2008). *Robotics and science literacy: thinking skills, science process skills and systems understanding.* Journal of Research in Science Teaching, 45(3), 373–394.

Toh, L. P. E., Causo, A., Tzuo, P. W., Chen, I. M., & Yeo, S. H. (2016). *A Review on the Use of Robots in Education and Young Children*. Educational Technology & Society, 19 (2), 148–163.

Williams, D. C., Ma, Y., Prejean, L., Ford, M. J., & Lai, G. (2007). *Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp.* Journal Research on Technology in Education, 40(2), 201-216.

Young, S. S. C., Wang, Y. H., & Jang, J. S. R. (2010). *Exploring perceptions of integrating tangible learning companions in learning English conversation*. British Journal of Educational Technology, 41(5), E78-E83.

7. APPENDIX

Robotic Interventions for improving educational outcomes - PROTOCOL

Description of the condition - Description of the intervention

In recent years, teachers are increasingly trying to include robotic activities in teaching process due to the benefits they seem to have. The application of robotics in education has yielded significant positive results in the development of technology literacy and the ability to problem solve (Anagnostakis & Michaelides, 2006). Robotics make learning fun and sharpen student's thinking. Sullivan (2008) suggests that student's engagement with robotics, and more particular with activities that require observation, designing, calculation, measurements and the ability to test their projects in real life context, increase their imagination. She also describes better intuitive assessment of hypotheses and variables, improved skills in observation, calculation, construction and creativity.

Studies indicate equally positive results in social skills. Children develop the spirit of team work and mutual collaboration (Mitnik et al, 2008, Nuget et al, 2009, Owens et al, 2008).There are different approaches on how robotics is used in educational setting. It can be used as a cognitive tool in a particular lesson, as a teaching technique, or as a subject area itself. We study robotics as a cognitive tool assisting a subject. The most popular subjects that are more likely to adopt these kind of activities is STEM (Science, Technology, Engineering and Mathematics). Barak & Zadok (2009) suggest that the benefits of educational robotics are relevant to concepts in Physics and Technology. Bellou & Mikropoulos (2013) have also proposed educational robotics as mind tools in Physics and Computer Science education through meaningful learning activities.

How the intervention might work

There has been an increasing body of research that reports positive outcomes from applying educational robotics in traditional classroom. Educational theorists believe that robotic activities have tremendous potential to improve classroom teaching (Papert 1993).

Some results indicate increased mean scores from pre and post - tests in science, technology and engineering concepts (Toh et al., 2016). Also, Barker & Ansorge's study (2007) reveals that the use of robotics in education help children improve their scores and develop academic skills (such as mathematical thinking and science process understanding). Results from another study (Kazakoff et al., 2013) showed that robotic and programming workshops help children in early childhood increase their sequencing skills. According to Barack & Zadok (2009) children's involvement with robotics make them come up with more inventive solutions to a problem.

In addition there are positive results in children's social skills as robotics reinforce collaboration (Toh et. al, 2016).

Why it is important to do this review

There are several benefits associated with the use of robotics in a curriculum and there is a rapid development of using technology and multi-media tools in education. Robotics is a trend and they become increasingly popular for educational purposes. Teachers all over the world develop activities and share their ideas to embody robotics into the teaching progress. Theoretical studies report robotics' benefits in academic performance. However, despite the increasing use of robotics, only a few studies focus on the investigation of the impact in an empirical way, so there is lack of empirical data In order to answer these questions and understand the association between use of robotics and increased academic performance we performed this systematic review.

Objectives

1) To assess the effectiveness of using robotic interventions as an educational tool in improving educational outcomes in K-12 students.

2) To explore the circumstances under which robotic interventions are more effective

Methods

Criteria for considering studies for this review Q

Types of studies

Only studies that presented a quantitative assessment of the benefits of robotics in learning were considered.

Experimental (randomized) & quasi-experimental studies that include comparison with a control group and provide a quantitative assessment of the benefits of robotics in learning.

Only studies conducted after 2000 will be included due to the rise of the interest of the academic community.

Types of participants

The participants should be elementary, middle and high school students (K-12).

Types of interventions

Articles that include interventions that use robotics as a teaching tool for a certain subject. In other words, the objective is not to teach robotics (such as in robotic courses) but using robotics as an educational means to teach another subject. This subject could be mathematics, physic, engineering etc.

Types of outcome measures

Primary outcomes: Achievement scores in major courses (mathematics, physics, language etc), problem-solving abilities and sequencing skills

Secondary outcomes : Team skills and collaboration, withdrawal and satisfaction

Search methods for identification of studies

Searches will be based on the following search string:

((teaching OR learning OR teach OR learn OR education OR educational) AND (robotic OR robotics OR robot OR robots OR Lego) AND (school OR k-12)).

It will be adapted for other databases using appropriate controlled vocabulary and syntax.

Electronic searches

The following databases will be searched:

(a) IEEE XPLORE,

(b) ACM Digital Library,

(c) ScienceDirect,

(d) SpringerLink,

(e) ERIC (Educational Resources Information Center)

(f) Wilson Education,

(g) Scopus

(h) MIT

Searching other resources

Grey literature

We will search relative journals to identify relevant unpublished studies and ongoing trials and Conference Proceedings Citation Index to identify conference proceedings.

We will search the ProQuest Dissertations & Theses Database to identify dissertations, theses and other unpublished literature.

Reference lists

We will search the reference lists of relevant review articles and included and excluded studies to identify additional studies in the published or unpublished literature.

Data collection and analysis

Selection of studies

Titles and abstracts from the literature searches will be independently screened by two review authors (Athanasiou and Mikropoulos). For eligible studies, full papers will be retrieved and judged independently by the authors to identify those satisfying the inclusion criteria. In case of uncertainty or disagreement, concurrence will be resolved through discussion or by consulting a third assessor (Mavridis) in order to reach a consensus about the study's eligibility.

Data extraction and management

For eligible studies, two review authors will extract the data using data extraction forms. We will extract data concerning:

- General information: title, published/unpublished, authors, year of publication, country, date of data extraction, sponsors
- Participants: sample size/number or participants randomized to the study, number of withdrawals, age/grade, gender, ethnicity, socioeconomic status
- Intervention: type(s), robot types used in the research, duration and intensity of intervention,
- Outcome: outcome measures, the subject that the researcher wanted to teach through robotics,, primary/secondary outcomes, effect sizes

We will resolve discrepancies through discussion or, if required, we will consult a third person (Mavridis). When information regarding any of the above is unclear, we will attempt to contact the corresponding authors of the original reports to provide further details. We will present the data in the "Characteristics of Included Studies Table".

Assessment of risk of bias in included studies

Two review authors will code each included study using the Cochrane tool for assessing risk of bias (Higgins 2011). This includes the assessment of selection bias (random sequence allocation and allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment and incomplete outcome data), reporting bias (selective outcome reporting) and other sources of bias. We will report any study characteristic that seems peculiar and may affect the magnitude of effect for the robotic interventions. The review authors will judge the risk of bias as either 'high risk' of bias, 'low risk' of bias or 'unclear risk' of bias, with information provided that led to each judgement. We will resolve disagreements by discussion and, if necessary, we will contact the third assessor for advice. We will provide details on each study's risk of bias in a 'Risk of bias' table.

Measures of treatment effect

We will calculate overall effects from the trials where data are available.

Continuous (including scale) data

We will analyze continuous outcomes using the mean difference (MD) with 95% confidence intervals (CI). Mean differences and 95% CI will be calculated comparing and pooling the mean score differences from the end of treatment to baseline for each group. If studies use different scales to measure outcome we will report Standardized Mean Differences (SMD) and 95% confidence intervals

Dichotomous (binary) data

We will analyze dichotomous outcomes calculating the Relative Risk (RR) for each trial with the uncertainty in each result being expressed by their 95% CI.

Multiple measure strategies

Investigators may choose different instruments to measure outcomes, either because they use different definitions of a particular outcome or because they choose different instruments to measure the same outcome. We aim to use the standardized mean difference to combine trials that assess the same outcome with different measures or instruments. If there is evidence of skewed data, this will be reported. In case of missing data about the standard deviation of the change we will impute this measure using the standard deviation at the end of treatment for each group.

Unit of analysis issues

Cluster- randomized trials

Cluster- randomized trials randomize groups of people rather than individuals. For each cluster- randomized trial, we will first determine whether or not its data incorporate sufficient controls for clustering (such as robust standard errors or hierarchical linear models). If the data do not have proper controls, then we will attempt to obtain an appropriate estimate of the data's intra-cluster correlation coefficient (ICC). If we cannot find an estimate in the report of the trial, then we will request an estimate from the trial report authors. If the authors do not provide an estimate, then we will obtain one from a similar study and conduct a sensitivity analysis to determine if the results are robust when different values are imputed. We will use the ICC estimate to control the trial's data for clustering, according to procedures described in Higgins 2011. This process will prevent meta-analyses from being based on clustered data that have not been properly controlled.

Cross-over trials

We will consider cross-over trials as eligible for inclusion if participants were randomized into the first period. For cross-over trials, we will extract and analyze data from the first period only.

Dealing with missing data

Where data are not reported for some outcomes or groups, we will attempt to contact the trial authors to request missing data and further information on dropouts. If the original investigators cannot provide the missing data, we will assume these data are not missing at random and that missing cases had poor outcomes. We will report missing data, dropouts, and reasons for dropout, and discuss the potential impact of this missing information on the findings of this review.

Assessment of heterogeneity

Differences among included studies are discussed in terms of their participants, interventions, outcomes, and methods. For each meta-analysis, we will use three methods to identify statistical heterogeneity: visually inspecting forest plots to see if the confidence intervals of individual studies have poor overlap; conducting a Chi2 test; and calculating an I2 statistic. We will consider a meta-analysis to have heterogeneity if its Chi2 P value is less than 0.10 and its I2 statistic is greater than 50%. Furthermore, we will assess the cause of heterogeneity by conducting subgroup and sensitivity analyses.

Assessment of reporting biases

We will aim to minimize publication bias by searching electronic databases and hand searching, grey literature including conference abstracts.

We will be alert to possible duplication bias by crosschecking details of authors, locations, numbers of participants and dates.

If we identify a sufficient number of studies (at least 10) we will create a funnel plot to evaluate the association between effect size and standard error. An asymmetric plot may indicate publication bias (small-study effect or reporting bias). In that case we will do a sensitivity analysis that indicates whether changing the weight given to small studies changes the results of the analysis.

Data synthesis

We will use Stata12 to conduct all meta-analyses. All meta-analyses will be conducted using the random-effects model. Relative risks or rate ratios and 95% confidence intervals (CIs) will be calculated for dichotomous outcomes and combined using Mantel-Haenszel methods.

Subgroup analysis and investigation of heterogeneity

Provided that we have at least 10 studies we plan to do the following subgroup analyses, and to do a Chi2 test for each analysis to determine whether or not the effects of robotics are statistically significantly different for different subgroups:

- Age (primary/elementary school vs. secondary/high school).
- Gender (male vs. female).
- Intervention duration (< 1 month vs. ≥ 1 month).

Sensitivity analysis

We will investigate the influence of study characteristics on the robustness of the review results by conducting sensitivity analyses. More specifically, we will conduct the following sensitivity analyses

- Exclusion of studies at high/unclear risk of bias for allocation concealment and sequence generation

Protocol references

1. Anagnostakis, S., Michaelides, P.G. (2006): ‘Laboratory of Educational Robotics’ - An undergraduate course for Primary Education Teacher – Students. In: 3rd International Conference on Hands-on Science, pp. 329—335. Braga, Portugal

2. Barak, M., & Zadok, Y. (2009). Robotics projects and learning concepts in science, technology and problem solving. International Journal Technology & Design Education, 19(3), 289-307.

3. Barker, B. S., & Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. Journal Research on Technology in Education, 39(3), 229-243.

4. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, 5.1.0. [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

5. Kazakoff, E. R., Sullivan, A., & Bers, M. U. (2013). The Effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. Early Childhood Educational Journal, 41, 245–255.

6. Mikropoulos, T.A., Bellou, I. (2013): Educational Robotics as Mindtools. Themes in Science and Technology Education. 6(1), 5--14. 12.Papert, S.: Mindstorms. Children, Computers and Powerful Ideas. Basic Books, New York (1980)

7. Mitnik, Ruben, Nussbaum, M., & Soto, A. (2008): An autonomous educational mobile robot mediator. Autonomous Robots. 25(4), 367--382

8. Nuget, G., Barker, B., Grandgenett, N., Adamchuk, V. (2009): The Use of digital manipulatives in K-12: robotics, GPS/GIS and programming. In Atman, C. (ed.) Proceedings of the 39th IEEE international conference on Frontiers in education conference, pp. 302--307. IEEE Press, USA

9. Owens, G., Granader, Y., Humphrey, A., Baron-Cohen, S. (2008): LEGO® Therapy and the Social Use of Language Programme: An Evaluation of Two Social Skills Interventions for Children with High Functioning Autism and Asperger Syndrome. J. Autism and Developmental Disorders. 38(10), 1944--1957

10. Papert, S. (1993). Mindstorms: Children, computers, and powerful ideas (2nd ed.). New York, NY: Basic Books.

11. Sullivan, F.R. (2008): Robotics and science literacy: Thinking skills, science process skills and systems understanding. J. Research in Science Teaching. 45(3), 373--394

12. Toh, L. P. E., Causo, A., Tzuo, P. W., Chen, I. M., & Yeo, S. H. (2016). A Review on the Use of Robots in Education and Young Children. Educational Technology & Society, 19 (2), 148–163.