



Systematic literature review: teaching novices programming using robots

L. Major T. Kyriacou O.P. Brereton

School of Computing and Mathematics, Keele University, Staffordshire, UK
E-mail: l.major@epsam.keele.ac.uk

Abstract: Teaching programming to novices is a difficult task due to the complex nature of the subject, the negative stereotypes are associated with programming and because introductory programming courses often fail to encourage student understanding. This study investigates the effectiveness of using robots as tools in the teaching of introductory programming and to determine whether such technology can help to overcome the current barriers for learners in this context. The systematic literature review (SLR) methodology is used to address this aim. Nine electronic databases, the proceedings from six conferences and two journals were searched for relevant literature and exclusion criteria, and after performing several validation exercises, in total, 75% of included papers report that robots are an effective teaching tool and can help novice programmers in their studies. Most of these papers focus on the use of physical robots, however, and further research is needed to assess the effectiveness of using simulated robots.

1 Introduction

Learning to program a computer has long been recognised as a difficult task for novices [1]. This has resulted in introductory programming courses suffering high drop-out rates [2] and many first-time programmers making little progress in their studies [3]. Programming is also associated with several negative stereotypes. These include the misconceptions that programming is so complex that most novices will never be able to become competent programmers [4] and that learning to program is rarely anything but a solitary and uninspiring experience [5].

Various efforts have been made by educators to try and overcome the difficulties that novice programmers encounter. Such attempts are often referred to as interventions. The work that is presented here focuses upon the use of robots as teaching tools in order to teach programming. The study of such an intervention was chosen at the School of Computing and Mathematics in Keele University, which has some experience in using robotics. In addition, a preliminary analysis of the literature demonstrated how robots can help students to better understand the algorithms that they have created [6].

This study employs the systematic literature review (SLR) methodology [7] to investigate the use of robots as tools to aid the process of teaching programming. The SLR is a trustworthy, rigorous and auditable tool [8] and is one that allows for existing evidence to be collected and summarised while identifying gaps in current research [7]. No past SLR has been found to examine the use of robotics in such a context.

The SLR methodology that has been implemented is described in depth in Section 2, while Section 3 is

dedicated to the results of the SLR search. In Section 4, a discussion takes place in an attempt to answer six research questions and in regards to different aspects of the review (including details of validation exercises that have been undertaken and an expanded discussion which considers the implications of this research). This is followed by a conclusion in Section 5.

A short version of this paper was presented at the EASE 2011 Conference that was held at Durham University, UK [9]. This paper reports on additional research that has been undertaken to validate the findings of the original work. This has involved the removal of low-scoring papers from the aggregation in order to determine what effect this has upon the overall results of the SLR as well as an expanded discussion section. Use of the 'Snowball' method has also been adopted in order to further validate the results of the SLR. This was done by revisiting included literature and by analysing the 'Background' or 'Introduction' sections of these papers for references of interest that may have been overlooked during the initial search. Analysis of the publication source of literature included in the SLR has also been performed. The implications arising from performing these additional activities will be considered and discussed and recommendations for potential future research will be provided.

2 Method

2.1 Research questions

This report is based upon the SLR guidelines as proposed by Kitchenham and Charters [7]. A protocol was developed as part of the SLR [10]. The aim of the SLR was to determine

how effective the use of robotics has been in the teaching of introductory programming. Six research questions were created in order to achieve this goal:

- [RQ1]: What computer languages are being taught in introductory programming courses that make use of robots as teaching tools?
- [RQ2]: Are the robots that are being used simulated or physical?
- [RQ3]: What are the characteristics (i.e. what is the age, level of education, etc.) of the novices being taught?
- [RQ4]: What types of studies are being performed by researchers that investigate the teaching of introductory programming concepts using robots?
- [RQ5]: What is the scale (e.g. number of participants) of studies that are being performed by researchers?
- [RQ6]: Do collected studies suggest that using robotics to teach introductory programming is effective?

2.2 Search process

The search process comprised of manual and automatic searches of electronic resources. This strategy was deemed suitable after trial searches were performed when devising the SLR protocol. Nine electronic databases were searched during the SLR with five of these having been identified as potentially useful by a previous study [11]. The electronic databases searched were: ACM Digital Library, CiteSeerX, EBSCOhost, ERIC, IEEEExplore, ISI Web of Science, Keele University's Digital Library, Australian Education Index and British Education Index. Manual searches of conference proceedings also took place and six conferences were identified as potentially relevant: European Computer Science Summit (ECSS), European Symposium on Programming (ESOP), International Conference on Software Engineering (ICSE), Conference on Innovation and Technology in Computer Science Education (ITiCSE), Technical Symposium on Computer Science Education (SIGCSE) and Conference on Information Technology Education (SIGITE). The Journal of Information Technology Education and the Oxford Computer Journal were also examined to see if they contained any literature relevant to the study.

The following search strings were created in order to retrieve information from the electronic resources discussed. These search strings were formed after analysing the keywords of relevant literature that was found during a general search of the resources outlined above:

- (robots OR robotics) AND ('amateur programming' OR 'amateur programmer').
- (robots OR robotics) AND ('beginner programming' OR 'beginner programmer').
- (robots OR robotics) AND ('first time programming' OR 'first time programmer').
- (robots OR robotics) AND ('introductory programming' OR 'introductory programmer').
- (robots OR robotics) AND ('novice programming' OR 'novice programmer').
- (robots OR robotics) AND 'teaching programming'.
- (robots OR robotics) AND 'learning programming'.

The following search string was then used to search on the title and abstracts of papers alone:

- (robots OR robotics) AND programming AND (novice OR beginner OR introductory OR teaching OR learning OR CS1 or 'first time').

Use of a two-stage search method was chosen in order to ensure that all relevant material had been collected and to make searches of the electronic resources more manageable. A trial search was used to validate the effectiveness of the search strategy and three papers previously identified as relevant (after the general search of the literature) were returned during this [6, 12, 13].

2.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria were used to ensure that only relevant literature was accepted into the SLR.

2.3.1 Inclusion criteria:

1. Publications were only included that reported on the use of robotics in teaching introductory programming to students who were studying a specific computing or IT-related course.
2. Papers that involve an empirical study or have a 'lessons learned' (experience report) element were included.
3. Although several papers reported the same study, only the most recent paper was included.
4. Date of publication did not act as a barrier for inclusion.
5. Grey literature (such as technical papers or government reports) was accepted if relevant.

2.3.2 Exclusion criteria:

1. Publications were excluded if their main focus was not on the use of robotics in teaching computing or IT students introductory programming but on the use of robots in general education courses, as part of a non-IT or computing-related course syllabus or to teach rudimentary programming concepts to very young children.
2. Papers that just propose an approach or describe the use of robots to teach introductory programming (with no 'lessons learnt' component) were excluded.
3. Papers and reports were excluded when only the abstract but not the full text was available.
4. Publications were excluded if they are not written in English.
5. Letters, editorials and position papers were all excluded.

Section 3.4 considers the potential impact of adopting some of these inclusion and exclusion criteria upon the validity of the SLR.

2.4 Quality assessment

Each publication in the final set was assessed for its quality. This quality assessment procedure was performed during the data extraction phase and ensured that included studies made a valuable contribution to the SLR. The 11 criteria for quality assessment are discussed by Dybå and Dingsøy [14]. These criteria were used in an SLR when there were a number of different study types. Use of the same criteria was deemed appropriate during this SLR as it was envisaged that it would also include studies of several sorts.

The 11 criteria used to assess the quality of each publication were:

1. Is the paper based on research or is it a ‘lessons learned’ report based on expert opinion?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to address the aims of the research?
5. Was the recruitment strategy appropriate to the aims of the research?
6. Was there a control group with which to compare treatments?
7. Was the data collected in a way that addressed the research issue?
8. Was the data analysis sufficiently rigorous?
9. Has the relationship between researcher and participants been considered to an adequate degree?
10. Is there a clear statement of findings?
11. Is the study of value for research or practice?

The first two of these criteria were used to exclude non-research papers and those that did not clearly state the aims of their research. This represents the minimum quality threshold that was observed during this SLR. The remaining nine criteria are aimed in determining the rigour and credibility of the research methods employed as well as the relevance of each paper to the SLR. The answers to each question (in regard to each item of literature included in the SLR) were tabulated and assigned a value of 1 (‘Yes’) or 0 (‘No’). In order to test the validity of the quality assessment procedure, a second reviewer (TK) was given a random sample of seven papers and was asked to assess their quality based on the same quality assessment criteria outlined. There was no disagreement on the overall quality assessment of these papers.

2.5 Data extraction

In order to answer the research questions discussed in Section 2.1, the following data were extracted from each study included in the SLR:

- Abstract and bibliographic reference;
- Why the study was accepted into the SLR?
- Study type (e.g. journal paper, conference paper);
- Study aims and objectives;
- Setting of the study;
- Methodology of study (e.g. observational, experience report, comparative);
- Information about baseline where appropriate (i.e. method against which robotics is being compared);
- Number of participants in a study (e.g. number of students in an experiment);
- How data was collected and analysed during the study;
- Characteristics of the novices being taught (e.g. age, level of education);
- Type of computer language being taught using robots (e.g. Java, C++, others);
- Nature of the robot being used to teach the programming language (e.g. simulated or physical);
- Findings and conclusions;
- Relevance of the study (e.g. in relation to the topic under consideration);

- Effectiveness of robotics as an intervention to teach introductory programming;
- The study quality assessment.

All data were extracted by a sole reviewer (LM) while the second reviewer (TK) independently extracted information from a random sample of seven publications. These results were then compared. As no significant anomalies were evident from this validation activity, the data extraction strategy was deemed to be appropriate. All extracted data were stored in a spread-sheet.

3 Results

This section summarises the results of the study.

3.1 Search results

After reading the full text of articles that were returned as a result of the search process, 34 studies were deemed to be relevant to the SLR and were accepted into it. Before this figure was arrived at 60 papers were read in their entirety. Of these papers 26 were considered to be either irrelevant or incompatible with the inclusion criteria. A manual analysis of relevant conferences and journals that took place alongside the automatic searches of the electronic databases and one study was included as a result. Appendix 1 lists all the articles included in the review, whereas Appendix 2 displays the results of the automatic search process. Fig. 1 shows the year of publication for studies accepted into the SLR. Only one paper published before 2000 has been included despite the inclusion and exclusion criteria placing no bar upon the year of publication.

3.2 Quality assessment of included studies

In Section 2.4 the quality assessment strategy used during the SLR is discussed. The results of this quality evaluation are

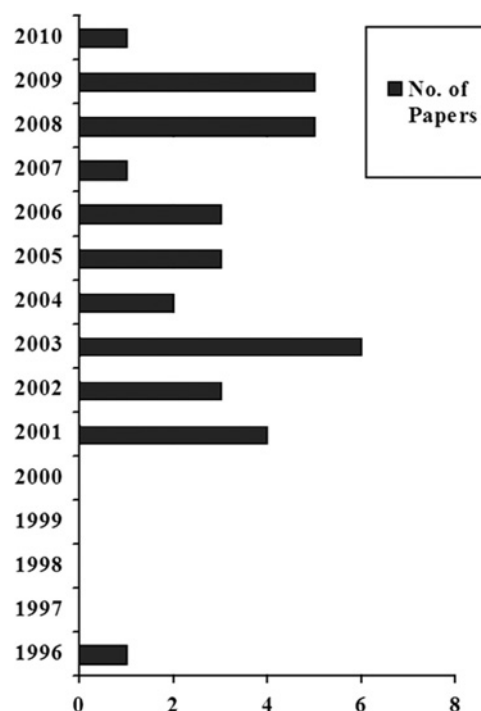


Fig. 1 Publication year of included studies

presented in Table 1 and each paper has been assigned a quality score out of 11. All articles included in the review were based on research or presented a 'lessons learned account' and clearly stated the aims of the research. Of the 34 studies 27 offered some description of the context in which the research was carried out while 25 were considered to have an appropriate research design.

Analysis of Table 1 also displays how many of the studies included had an inadequate recruitment strategy, failed to use a control group, did not collect (or sufficiently analyse) data in a way that addressed the research issue and did not consider the relationship between participants and the researcher. The majority of studies that scored 0 in respect to these criteria are examples of studies that offered a 'lessons learned' account and did not report any empirical data. Three of the studies included in the review were awarded the maximum score of 11. The lowest score that articles were awarded was a three. The average quality score of the papers included in the review is 6.9. The median score of papers included in the SLR is six (with 11 studies awarded this score).

As the average quality score of included articles varied widely, it was decided that it would be interesting to remove

low-scoring papers from the aggregation and to analyse the effect that such a change would have. Nine papers were awarded a quality score of five or less during the first round of quality assessment (as documented in Table 1). These papers were removed from the aggregation during the second round of quality checking. When the nine low-scoring papers are omitted from the aggregation, the average quality score of included literature rises from 6.9 to 8 (out of 11). The implications of removing low-scoring papers from the set of included studies, in regards to the effectiveness of using robotics to teach introductory programming, are discussed in further detail in Section 3.3 [RQ6].

3.3 Research questions

Answers to the research questions outlined in Section 2.1 will now be discussed. A summary of the information that has been extracted from each included paper can be found in Appendix 3.

[RQ1] What computer languages are being taught in introductory programming courses that make use of robots as teaching tools?

Table 1 Results of SLR quality evaluation

Study	Research	Aim	Context	Research design	Recruit. strategy	Control group	Data collection	Data analysis	Relations-hip	Findings	Value	Total
[Barnes02]	1	1	0	0	0	0	0	0	0	0	1	3
[Becker01]	1	1	1	1	0	0	0	0	0	1	1	6
[Bell08]	1	1	1	1	0	0	0	0	0	1	1	6
[Borge04]	1	1	1	1	1	0	1	1	0	1	1	9
[Brauner10]	1	1	1	1	1	0	1	1	1	1	1	10
[Buck01]	1	1	0	0	0	0	0	0	0	1	1	4
[Cliburn06]	1	1	1	1	0	0	0	0	0	1	1	6
[Czejdo09]	1	1	0	0	0	0	0	0	0	0	1	3
[Enderle08]	1	1	1	1	0	0	0	0	0	1	1	6
[Fagin01]	1	1	0	0	0	0	0	0	0	1	1	4
[Fagin03a]	1	1	1	1	1	1	1	1	1	1	1	11
[Fagin03b]	1	1	1	0	0	0	0	0	0	1	1	5
[Flowers02]	1	1	1	1	0	0	0	0	0	1	1	6
[Garrett05]	1	1	1	1	0	0	0	0	0	1	1	6
[Goldweber01]	1	1	0	0	0	0	0	0	0	0	1	3
[Hirst03]	1	1	1	1	0	0	0	0	0	1	1	6
[Imberman07]	1	1	1	1	1	0	1	1	0	1	1	9
[Jadud03]	1	1	1	0	0	0	0	0	0	0	1	4
[Kurebay06]	1	1	1	1	1	0	1	1	0	1	1	9
[Ladd05]	1	1	1	1	0	0	0	0	0	1	1	6
[Lauwers09]	1	1	1	1	1	1	1	1	0	1	1	10
[Lawhead02]	1	1	1	0	0	0	0	0	0	0	1	4
[Lemone96]	1	1	1	0	0	0	1	0	0	1	1	6
[Mcwhorter09]	1	1	1	1	1	1	1	1	1	1	1	11
[Petre04]	1	1	1	1	1	0	1	1	0	1	1	9
[Sartzemi03]	1	1	0	1	1	0	1	0	0	1	1	7
[Sartzemi05]	1	1	1	1	0	0	0	0	0	1	1	6
[Sklar06]	1	1	1	1	1	0	1	1	0	1	1	9
[Summet09]	1	1	1	1	1	1	1	1	1	1	1	11
[Vandelden08]	1	1	1	1	0	0	0	0	0	1	1	6
[Weiss08]	1	1	0	1	0	0	0	0	0	1	1	5
[Wellman09]	1	1	1	1	1	1	1	1	0	1	1	10
[Williams03]	1	1	1	1	1	0	1	0	1	1	1	9
[Wu08]	1	1	1	1	1	1	1	1	0	1	1	10
total	34	34	27	25	14	6	15	12	5	29	34	/

When analysing the studies included in the SLR, 10 different categories were established in regard to the programming languages used. Java was the largest contributor to the SLR having been the main programming language used in 10 papers. This was followed by seven papers that reported on the use of a combination of programming languages. C++ was described by four papers contained in this work while the use of Not Quite C (NQC) and C was reported in one paper each. Ada (three papers), Python (two papers) and Scheme (one paper) have also been discussed. Evidence was also collected that highlights how attempts have been made to use specially designed programming languages in order to teach programming principles. This includes use of the Robolab software (two papers) in addition to the Scratch (one paper) and Dolittle (one paper) educational languages. The development of a customised language that integrated use of the Alice animation software was also reported by one study. Fig. 2 presents a summary of this information.

[RQ2] Are the robots that are being used simulated or physical?

Twenty-three of the 34 papers included in the SLR report the use of physical robots. In contrast, seven papers were included that discussed the use of simulated robots alone, whereas four papers reported on the implementation of both physical and simulated robotic technologies simultaneously. The 23 papers that report the use of a physical robot can be further divided to reflect the type of robot used. After analysis it was found that 14 papers describe the implementation of Lego Mindstorm technology while the iCreate and Scribbler robots were discussed in two papers each. One paper was found to detail the use of a custom robot, whereas four studies were found to make use of several types of robot at the same time. See Fig. 3 for a breakdown of this information in graphical form.

[RQ3] What are the characteristics of the novices being taught?

The different context of each study was scrutinised in order to determine the characteristics of the novices that have been taught introductory programming using robotics. Three different groupings were established as a result of this and these were university, high school or 'various'. Out of the 34 papers 23 reported on the use of robotic technology in a university setting, seven were based in a high school and four discussed the implementation of robots in several different environments.

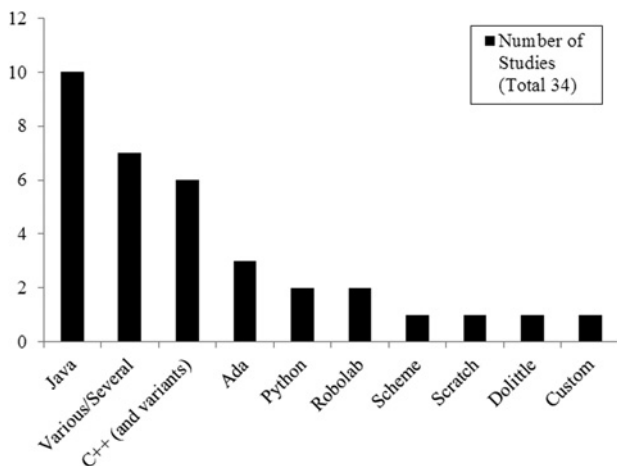


Fig. 2 Computer languages used by the papers included in the SLR

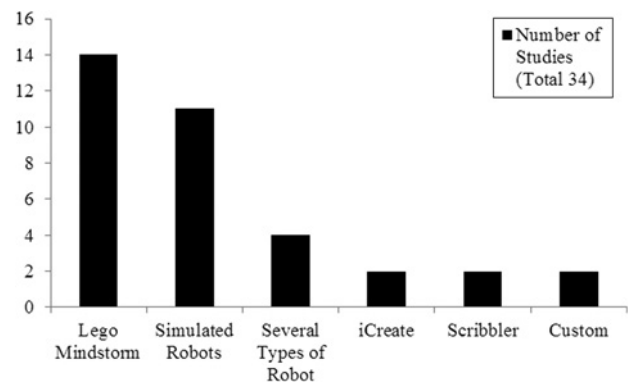


Fig. 3 Breakdown of type of robot used by papers included in the SLR

[RQ4] What types of studies are being performed by researchers that investigate the teaching of introductory programming concepts using robots?

The use of course critique surveys and questionnaires was the most commonly found method by which included studies evaluated their findings and proposals (nine papers reported the use of such methodologies). Interviews and focus groups have also been described (in two papers) as has the implementation of pilot lessons (two papers). In addition, comparative analysis has also taken place. This has included contrasting the effect on the learners of learning with robots to learning without (one paper) as well as a comparison of physical and simulated technologies (also one paper). While not studies as such, analysis of student grades has also been reported in two papers while one paper examined the impact that robots had upon retention rates. The remaining papers included in the review (16 papers) offered a 'lessons learned' or experience report account or did not explicitly state the results of any experiments undertaken.

[RQ5] What is the scale of studies that are being performed by researchers?

The 16 papers included in the review are examples of 'lessons learned' or experience style reports, whereas 18 papers offer evidence that an empirical study took place. The scale of studies included in the SLR varied widely. These ranged from small-scale studies which contained 15, 20 and 31 participants through to larger studies which reported on sample sizes of 121 and 151 students. One study compares the test results of hundreds of participants who took part in both robotics and non-robotics-based classes [12] and the potential implications of including such a large study in the review are discussed in further detail in Section 4.1. Only 12 papers report the exact number of students that took part in the research performed. In contrast six papers discuss conducting experiments or collecting information from participants but do not state the precise number of participants involved.

[RQ6] Do collected studies suggest that using robotics to teach introductory programming is effective?

After analysing all the papers included in the SLR, it is possible to present a breakdown on whether the included literature reports the use of robots to be an effective intervention when teaching introductory programming. Of the 34 papers included in the review, 25 papers report that the use of robotics is effective when teaching introductory programming concepts, five offer mixed results while only one paper states that robots were found to be ineffective.

Three papers were determined to be unclassifiable and do not provide a measure of the effectiveness of robots when used in such a context. Fig. 4 displays this information in the form of a pie chart.

It is possible to expand upon these findings by comparing these results with those that were found in response to RQ2 ('Are the robots that are being used simulated or physical?'). Of the 23 papers that report the implementation of a physical robot, 16 of these describe this as being effective, four papers offer a mixed verdict, two studies were unclassifiable, while one paper concluded that such a technique was ineffective. Of the seven papers that examined the use of simulated robots, six found such an intervention to be effective while one paper was unclassifiable. Of the four papers that describe the use of both physical and simulated robot technology, three papers suggest robotics to be effective while one study offers a mixed verdict. A breakdown of the effectiveness of robots as tools to teach programming, arranged by robot type, can be seen in Fig. 5.

In regards to the effectiveness of using robots as programming teaching tools, it was decided that it would be beneficial to look at the effect of removing the nine papers with a quality assessment score of five or less. This is because past research has found how low-quality studies reported significantly larger effects (i.e. impact of treatment) than good-quality studies [15]. This can alter the interpretation of benefit in regards to a particular intervention. Other work has also been found to support the theory that low-quality studies show more beneficial treatment effects than high-quality trials [16].

Nine papers were excluded from this analysis as they had been awarded a quality score of five or below. This left a remaining subset of 25 papers. A breakdown of the effectiveness of robots as tools to teach programming,

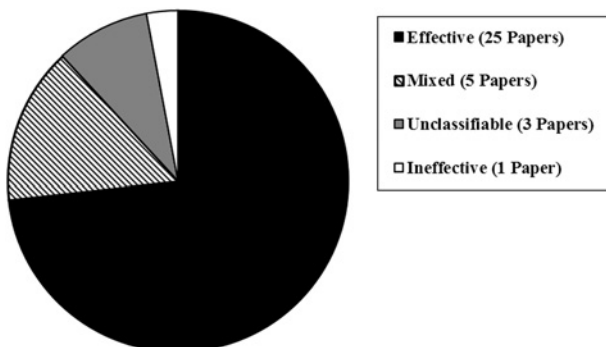


Fig. 4 Effectiveness of robots as tools to teach programming

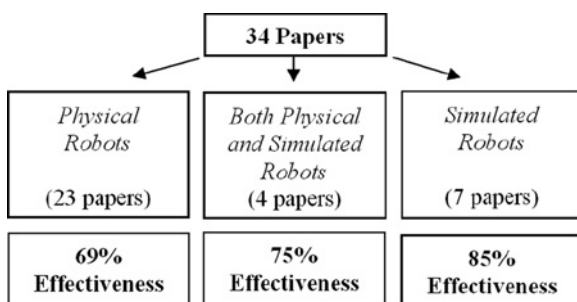


Fig. 5 Breakdown of effectiveness of robots as tools to teach programming (by nature of robot reported on)

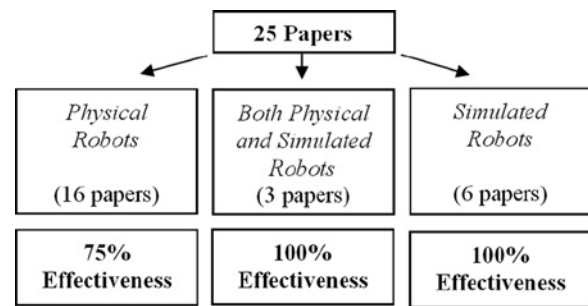


Fig. 6 Breakdown of effectiveness of robots as tools to teach programming after excluding those papers awarded a quality score of five or less

arranged by robot type and with low-scoring papers excluded from consideration, is displayed in Fig. 6.

Several points can be noted as a result of removing the low-scoring papers included in the SLR, in regards to the effectiveness of using robots as tools to teach programming. First, those papers which discuss the use of simulated robots appear largely unaffected by the decision to exclude papers that did not score well in the quality assessment. This is in contrast to those papers that discuss the use of physical robots as seven of these have been removed. It must be considered that more papers in the original set of 34 discussed the use of physical robots, however, and that substantially fewer papers examined the implementation of both simulated and physical robots together or just simulated robots alone. Also, as a result of removing low-scoring papers, it is interesting to note that five of these offered either mixed or unclassifiable results. Therefore the removal of the low-scoring papers offers greater evidence that supports the hypothesis that robots are an effective intervention when used in the teaching of introductory programming. This is because fewer papers offer mixed or unclassifiable results and could indicate that the research design adopted during these studies was not as rigorous as those used in some of the high-scoring studies. Of the 25 papers awarded a quality score of six or more, 12 of the 16 papers which discuss the use of physical robots offer positive findings in regards to their effectiveness while all six of the papers which discuss the implementation of simulated robots state that the use of such a method was effective. The potential implications that these findings may have, in addition to a consideration of other factors, is discussed in Section 4.3.

3.4 Limitations of SLR

The main threats to the validity of SLR are in relation to bias in the selection of publications and inaccurate data extracted. Search strings were devised as the review employed mainly electronic resources. These were developed after implementing trial searches, consulting experts and using a thesaurus. Despite this it is not possible to guarantee that all studies relevant to the topic under consideration were returned and there is a slight risk that some studies may have been omitted due to the search terms used (this will be discussed in greater depth in Section 4.1). Moreover, publication bias (the phenomena where 'negative' results are less likely to be published) may also have had some impact on the findings of the SLR although it is difficult to ascertain whether this was the case. The data extraction process may have also been negatively impacted by bias

when selecting articles. This is because the data extraction procedure was performed by one reviewer. The development of a SLR protocol and the use of a sample quality checking strategy (by a second reviewer) help to ensure that this was not the case. Finally, it is possible that the inclusion criteria may have inadvertently excluded some relevant resources. This is because the implemented criteria barred papers that contained no 'lessons learned' component or were related to the teaching of very young children. In addition, non-English language and abstract only papers were excluded from inclusion in the SLR. While no papers were excluded from this study due to the language they were written in or as only the abstract of the paper was available, such exclusion criteria could have inadvertently prevented valid work from being included in the SLR. As a result, it is recommended that researchers performing future studies, similar in scope to the one presented, consider the potential effects of adopting such inclusion criteria as it may have a detrimental impact upon their findings.

4 Discussion

This section presents details of additional research that has been undertaken to validate the findings of the SLR. In addition a discussion in regards to the results of the SLR is also described.

4.1 Additional validation of results (post-study)

To further validate the results of the SLR, additional validation activities have been undertaken after the completion of the study proper. These measures have been taken in addition to the validation strategies already outlined (such as a second reviewer independently extracting data from a random sample of papers to test the quality assessment and data extraction strategy).

It was decided that a suitable method of ensuring that all suitable literature had been collected was by adopting the 'Snowball' technique. This was done by revisiting the full set of 34 papers included in the review and by examining the 'Background' or 'Introduction' sections of those papers for potentially relevant references. In addition, references retrieved from articles accepted into the SLR were also analysed for literature of interest that may have been overlooked during the initial search. It was decided beforehand that if four or more papers (around 12.5% of the original set) were found as a result of this validation exercise, and that these papers met the SLR inclusion criteria and therefore should have been included in the review, then it could indicate that the SLR search strategy implemented was critically flawed.

In some cases literature was collected and read after analysing the background sections of included papers although often such work was deemed irrelevant. Several times papers were read but were excluded from the review after further analysis. In total, it was found that two papers could have been included in the review after completing the post-SLR validation exercise. The references of these two papers are [17, 18].

These two additional papers have been subject to the same procedures that were applied to the other papers included in the review. This includes applying the quality assessment and data extraction strategy previously outlined (in Section 2). After performing the quality assessment checks, and getting the results of this verified by a second reviewer

(TK), [17] was awarded a quality assessment score of nine whereas [18] was awarded a score of six. The data that were extracted from these two additional studies can be seen in Appendix 4. The data extracted from these two papers serve to strengthen the original findings of the SLR as both papers discuss how the use of robots can be effective when used in the teaching of introductory programming. Both [17, 18], however, discuss the implementation of physical robot technology rather than simulated robots.

As only two additional papers were found after implementing the post-SLR validation exercise, it is considered that the search strategy that was adopted during the review was sufficiently effective and rigorous. One of the papers [17] was probably not found during the initial search as it does not appear to be indexed by the electronic resources that have been used during the SLR. The second paper found when post-validating the results of the SLR, however, could have been identified during the SLR search as it is indexed in the ACM Digital Library (which was one of the electronic resources used). After the analysis of the title, abstract and keywords of [18], it was noted how the term 'programming' does not appear in any of these. Instead programming terms such as 'repetition, selection and the use of basic functions' are referred to. As 'programming' was used in every search term that was used during the SLR, it is probable that this particular article was overlooked initially as it was not returned during any of the automatic searches. It is considered that this is not a weakness of the search terms themselves, as it is not unreasonable to presume that if a paper discusses the teaching of introductory programming then the term 'programming' would appear in either the title or abstract of the paper. Indeed, this is the case with all of the other papers that have been included in the SLR. Nonetheless, it does highlight how some relevant papers can occasionally be overlooked when performing an SLR search and demonstrates how it is important that appropriate terms are used by authors in the titles, abstracts and keywords of their papers if secondary reviewers are to successfully locate them. It was also noted when analysing the background and reference sections of those papers included in the SLR that a large number of the same references were repeated in different papers. This serves to increase confidence in the results of the SLR as it likely indicates that all important work has been identified and included in the study. This can also be considered, therefore, to validate the search terms that were used during the SLR. When the two papers that were found as a result of the post-SLR validation exercise are included in the overall aggregation of results, the findings of the SLR are not significantly altered. There is a minor positive change when they are included, however, and 27 studies report the use of robots in the teaching of introductory programming concepts to be effective, five offer mixed results while only one paper states that robots were found to be ineffective. In total 36 papers were identified and included in the final set after performing the post-SLR validation.

4.2 Publication source of literature included in the SLR

The source of papers included in the SLR has been examined in order to identify if there are any prominent journals or conferences consistently publishing work related to the teaching of introductory programming using robots. It was

noted during this process that of the 36 papers included in the SLR (including the two found during the post-SLR validation) 17 were published in conference proceedings, 15 in journals and four in other sources (for example as technical reports). The references of included articles have been examined to determine if there are any 'outlier' sources (i.e. have any sources published a substantial number of articles included in the review). Two of these 'outliers' have been identified: The Journal of Computing in Small Colleges (which published eight articles that have been included in the review) and the SIGCSE Technical Symposium on Computer Science Education (which is an annual conference and was also found to have published eight articles). Between them these two sources contributed 16 of the 36 articles accepted in the SLR. Owing to the high number of relevant articles being published in these two venues, it is recommended that these sources may be candidates for a manual search if future work is to build upon this SLR. However, at this point it has been decided that this is not necessary as the post-SLR validation exercise discussed in Section 4.1 would most likely have uncovered any other relevant references not already found. The sources of the remaining 20 articles were varied and no other conference or journal publications were identified as contributing a substantial number of articles to the SLR.

4.3 Discussion on the results of the SLR

Various observations can be made as a result of the review. In regards to the original quality score the authors believe that this figure is low (with the average being 6.9/11). As Table 1 ('results of SLR quality evaluation') displays, a high proportion of papers contained in the initial set of 34 lacked vital experimental features like a control group, whereas the analysis of collected data was often considered to be of a poor standard. This is due to 16 of the initial 34 papers included in the review being 'lessons learned' or experience style reports. Such papers do not score well against the quality assessment criteria that have been used. Following this initial analysis, the set of data was reduced, so that only those papers awarded a quality score of 5 or more were considered, in order to ensure that the low-quality studies included in the SLR did not artificially inflate the reported effectiveness of robots used as programming teaching tools. However, as discussed in Section 3.3, removing those papers which scored poorly in the quality assessment was not found to have a detrimental impact upon the results of the SLR. If anything, removing the low-scoring papers may actually strengthen the argument that the use of robots can be an effective teaching tool when used in an introductory programming course. This is because five of the nine low-scoring papers offered mixed or unclassifiable results and may arguably not be as reliable as those papers which could be deemed to offer definitive findings. When these papers are removed from the aggregation the reported effectiveness of all three methods (the use of physical robots alone, the use of simulated robots alone and the use of both physical and simulated robots together) improves. It should be noted, however, that the results of the SLR cannot be said to offer conclusive evidence that one particular type of robot is more effective when used as a tool to teach programming. This is because of the relatively small sample sizes involved when low-scoring literature is removed. As a result, and due to included papers being found to use a wide range of methods to collect data, statistical analysis

has not been undertaken. Nonetheless, the results of the SLR are still considered as extremely valuable and provide several platforms upon which future research could build in order to further investigate the research area. Potential implications for future work are considered later in this discussion.

Only one large-scale comparative study was included in the SLR [12]. This paper reports the results of a year-long experiment which compared the results from over 800 students on identical tests from both robotics and non-robotics-based laboratory sessions. Traditionally such a large study may be considered to offer far more compelling evidence than the results of small non-comparative studies. However, as the computer language implemented in [12] was based on a scaled-down version of Ada, it has been considered that the results of this large-scale study are not necessarily more valuable than those papers which discuss studies that had fewer participants. This is because the more elaborate features of Ada cannot be executed on the RCX hardware [19] (which was used during the study) and as a result it is not possible to regard the use of such a language as offering conclusive evidence. This is due to a reduced version of Ada unlikely being as conceptually difficult for novices to learn as a full-scale object-oriented language such as Java would be. Interestingly, however, the large-scale comparative study reported in [12] was the only paper included in the SLR which reported completely negative results in regards to the effectiveness of using robots as programming teaching tools. The authors of the paper identify a range of reasons why this may have been the case and state how the use of a simulator for the robots programming system may have helped to overcome the issues encountered during the study.

Several research questions were created in order to determine the value of using robotics when teaching introductory programming and to provide a broad overview of the topic area. Various findings and trends, in regards to the teaching of introductory programming using robots, can be noted as a result. These include the observations that:

- The Java programming language is the one that has been most frequently adopted by educators.
- The use of physical robots is more commonly reported than the use of simulated robots.
- Course critique surveys and questionnaires are the most commonly reported methods used to evaluate the effectiveness of robotic interventions (where a primary study has taken place).
- The number of participants who have taken part in research to evaluate the value of robotics in teaching introductory programming varies greatly from study to study.

On the whole, the results of the SLR suggest that the use of robots can be an effective teaching tool when used in an introductory programming course. This is because three quarters of papers included in the SLR explicitly state that robots are valuable when used in such a manner. Potentially the most interesting finding to arise from the results of the SLR, however, is that the use of simulated robots may be more effective than physical robots when used as tools to teach programming. As no papers report simulated robots to be ineffective the use of simulated robots may potentially be just as, if not more, effective than physical robots. Such a hypothesis is further supported by the fact that when low-quality scoring papers were removed from the aggregation those which discuss the use of simulated

robots were largely unaffected (as only one paper was removed from the set and this was deemed to offer unclassifiable results anyhow). This is in contrast to those papers that discuss the use of physical robots as four of the papers which found physical robots to be effective were removed. It should be noted, however, that the use of physical robots by educators has been much more commonly reported to date and that fewer studies have been found to evaluate the use of simulated tools of this nature. Moreover, due to included papers using a wide range of methods to collect data, in conjunction with some of the samples sizes involved being small, statistical analysis techniques have not been utilised during this study and so these findings cannot be said to be statistically significant. As a result, further research is required in order to determine the true effectiveness of simulated robots that can be used to support the teaching of programming. Nonetheless, this work highlights how there is the potential for further research to build upon the body of existing knowledge documented by the SLR.

The SLR demonstrates a clear need for large-scale and high-quality research to be undertaken in order to determine the true effectiveness of robots as programming teaching tools. As a result of this study, it is possible to identify several areas that future research may seek to investigate. First, it is important that work is carried out in order to determine the true value of using simulated robots as programming teaching tools, as the results of the SLR highlight how potentially the use of simulated robots may be more effective than the use of physical robots. Owing to the relatively small sample sizes involved, however, such a hypothesis needs to be rigorously investigated and tested. A second theme that future researchers could also follow is to scrutinise the merits of using different types of programming languages, with robot technology, in order to teach the subject. Such research may seek to investigate whether one computer language in particular is better suited for use with robots. Similarly, a consideration of the benefits of using different types of robots (e.g. different variants of physical robots) would also be beneficial. Finally, an examination of the broader hypothesis that using robots as programming teaching tools is more effective than other non-robotic approaches would also be significant and could serve to inform future teaching practices.

In order to contribute to knowledge in this research area, some additional work based on these suggestions has been completed. Described in [20] a study involving 23 trainee High School ICT/Computer Science teachers, which reports on their experiences using a robot simulator to teach Java programming concepts, has been performed. It was found during this study how the implementation of a robot simulator was found to offer an enjoyable and effective method of teaching programming despite only moderately improving the trainee's confidence in their ability to teach programming. It is intended that this work will now be further built upon by considering the value of using a robot simulator to teach programming concepts to novice programming students.

5 Conclusions

This study has examined the effectiveness of using robotics to teach introductory programming by using the SLR methodology. After implementing the search strategy 34 papers were initially included in the SLR. Post-SLR

validation exercises were then performed, including the removal of low-scoring papers from the review, an expanded discussion and the implementation of the 'Snowball' method. A further two relevant papers were identified and included in the final set as a result of these measures. These papers were subject to a pre-determined data extraction and quality checking strategy. The results of the SLR indicate how the use of robots can be an effective teaching tool when used in introductory programming courses. Indeed, 75% of literature included in the review reported this to be the case.

Various findings and trends, in regards to the teaching of introductory programming using robots, have been noted as a result of the SLR. These include the discovery that the Java programming language is the one that has been most frequently adopted by educators who use robots as tools to teach programming, that the use of physical robots has been more commonly reported than the use of simulated robots, that course critique surveys and questionnaires are the most commonly reported methods used to evaluate the effectiveness of robotic interventions (where a primary study has taken place) and that the number of participants who have taken part in research to evaluate the value of robotics in teaching introductory programming varies greatly from study to study.

The most important finding of the SLR, however, is that there is a demonstrable need for large-scale and high-quality research to be undertaken in order to determine the true effectiveness of robots as programming teaching tools. In particular, there is scope to further investigate the effectiveness of using simulated robots as tools to teach programming. As has been described some additional work has already been completed, and will continue, in order to investigate such a topic. Other potential areas that new research may seek to investigate have also been outlined in the paper presented.

6 Acknowledgment

L. Major would like to thank Mark Turner and Thomas Neligwa (both of Keele University) for their constructive comments when reviewing an early version of the SLR.

7 References

- 1 Kelleher, C., Pausch, R.: 'Lowering the barriers to programming: a taxonomy of programming environments and languages for novice programmers', *ACM Comput. Surv.*, 2005, 37, (2), pp. 83–137
- 2 Ala-Mutka, K.: 'Problems in learning and teaching programming – a literature study for developing visualizations in the Codewitz-Minerva project'. Tampere University of Technology, 2004
- 3 Robins, A., Rountree, J., Rountree, N.: 'Learning and teaching programming: a review and discussion', *Comput. Sci. Educ.*, 2003, 13, (2), pp. 137–172
- 4 Turkle, S.: 'The second self – computers and the human spirit' (MIT Press, 2005, 20th Anniversary Edition)
- 5 Esteves, E., Antunes, R., Fonseca, B., Morgado, L., Martins, P.: 'Using second life in programming's communities of practice'. Proc. CRIW 2008, Omaha, NE, USA, September 2008, pp. 99–106
- 6 Flowers, T.R., Gossett, K.A.: 'Teaching problem solving, computing, and information technology with robots', *J. Comput. Small Coll.*, 2002, 17, (6), pp. 45–55
- 7 Kitchenham, B., Charters, S.: 'Guidelines for performing systematic literature reviews in software engineering'. Technical report, Keele University and Durham University Joint Report, 2007
- 8 Kitchenham, B. Keele University: 'Procedures for undertaking systematic reviews'. Technical report, Keele University, 2004
- 9 Major, L., Kyriacou, T., Brereton, O.P.: 'Systematic literature review: teaching novices programming using robots'. Proc. 15th Int. Conf. on

- Evaluation and Assessment in Software Engineering (EASE 2011), Durham University, UK, April 2011, pp. 21–30
- 10 Major, L.: 'Teaching novices programming using robotics: a systematic literature review protocol', EBSE. http://www.dur.ac.uk/ebse/resources/studies/protocol/teaching_programming_with_robots_protocol_v1-4.pdf, accessed 10 February 2011
- 11 Brereton, P., Kitchenham, B., Budgen, D., Turner, M., Khalil, M.: 'Lessons from applying the systematic literature review process within the software engineering domain', *J. Syst. Softw.*, 2007, **80**, (4), pp. 571–583
- 12 Fagin, B.: 'Measuring the effectiveness of robots in teaching computer science'. Proc. 34th SIGCSE Technical Symp. on Science Education, Reno, Nevada, USA, February 2003, pp. 307–311
- 13 Petre, M., Price, B.: 'Using robotics to motivate 'back door' learning', *Educ. Inf. Technol.*, 2004, **9**, (2), pp. 147–158
- 14 Dybå, T., Dingsøy, T.: 'Empirical studies of agile software development: a systematic review', *Inf. Softw. Technol.*, 2008, **50**, (9–10), pp. 833–859
- 15 Moher, D., Jones, A., Cook, D.J., et al.: 'Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?', *Lancet*, 1998, **352**, pp. 609–613
- 16 Shang, A., Huwiler-Müntener, K., Nartney, L., et al.: 'Are the clinical effects of homeopathy placebo effects? Comparative study of placebo-controlled trials of homeopathy and allopathy', *Lancet*, 2005, **366**, (9487), pp. 726–732
- 17 Wong, K.: 'Teaching programming with LEGO RCX robots'. Proc. 18th Information Systems Education Conf., Cincinnati, Ohio, USA, November 2001
- 18 Imberman, S.P., Klibaner, R.: 'A robotics Lab for CS1', *J. Comput. Small Coll.*, 2005, **21**, (2), pp. 131–137
- 19 Fagin, B.: 'An ada interface for Lego Mindstorms', *Ada Lett.*, 2000, **20**, (3), pp. 20–40
- 20 Major, L., Kyriacou, T., Brereton, O.P.: 'Experiences of prospective high school teachers using a programming teaching tool'. Proc. 11th Koli Calling Int. Conf. Computing Education Research (Koli Calling '11), Koli National Park, Finland, 17–20 November 2011, pp. 126–131

8 Appendix 1

8.1 Articles included in the review

[Barnes02]: Barnes D.J.: 'Teaching Introductory Java through LEGO MINDSTORMS models'. Proc. 33rd SIGCSE Technical Symp. on Computer Science Education, New York, NY, USA, 2002, pp. 147–151

[Becker01]: Becker B.W.: 'Teaching CS1 with Karel the Robot in Java'. Proc. 32nd SIGCSE Technical Symp. on Computer Science Education, New York, NY, USA, 2001, pp. 50–54

[Bell08]: Bell S., Heeler L., Heeler P.: 'A preliminary report on the use of robots with elementary school students', *J. Comput. Small Coll.*, 2008, **23**, (4), pp. 263–268

[Borge04]: Borge R., Fjuk A., Groven A.K.: 'Using Karel J collaboratively to facilitate object-oriented learning'. Proc. IEEE Int. Conf. Advanced Learning Technologies, August 2004, pp. 580–584

[Brauner10]: Brauner P., Leonhardt T., Ziee M., Schroeder U.: 'The effect of tangible artefacts, gender and subjective technical competence on teaching programming to seventh graders'. Proc. 4th Int. Conf. Informatics in Secondary Schools – Evolution and Perspectives, Zurich, Switzerland, January 2010, pp. 61–71

[Buck01]: Buck D., Stucki D.J.: 'JKarelRobot: a case study in supporting levels of cognitive development in the computer science curriculum'. Proc. 32nd SIGCSE Technical Symp. on Computer Science Education, Charlotte, North Carolina, USA, February 2001, pp. 16–20

[Cliburn06]: Cliburn D.C.: 'Experiences with the LEGO Mindstorms throughout the Undergraduate Computer Science Curriculum'. Proc. Frontiers in Education Conference 36th Annual, San Diego, California, USA, October 2006, pp. 1–6

[Czejdo09]: Czejdo B.D., Bhattacharya S.: 'Programming robots with state diagrams', *J. Comput. Small Coll.*, 2009, **24**, (5), pp. 19–26

[Enderle08]: Enderle S.: 'Grape: graphical robot programming for beginners'. Proc. Research and Education in Robotics – EUROBOT, Heidelberg, Germany, 2008, pp. 180–192

[Fagin01]: Fagin B.S.: (2001) 'Teaching computer science with robotics using Ada/Mindstorms 2.0'. In Proc. 2001 Annual ACM SIGAda Int. Conf., pp. 73–78, ACM

[Fagin03a]: Fagin B.: 'Measuring the effectiveness of robots in teaching computer science'. Proc. 34th SIGCSE Technical Symp. on Computer Science Education, Reno, Nevada, USA, February 2003, pp. 307–311

[Fagin03b]: Fagin B.: 'Ada/Mindstorms 3.0', *IEEE Robot. Autom. Mag.*, 2003, **10**, (2), pp. 19–24

[Flowers02]: Flowers T.R., Gossett K.A.: 'Teaching problem solving, computing, and information technology with robots', *J. Comput. Small Coll.*, 2002, **17**, (6), pp. 45–55

[Garrett05]: Garrett A., Thornton D.: 'A web-based programming environment for LEGO Mindstorms robots'. Proc. 43rd Annual Southeast Regional Conference, Kennesaw, Georgia, Alabama, USA, March 2005, pp. 349–350

[Goldweber01]: Goldweber M., Congdon C., Fagin B., Hwang D., Klassner F.: 'The use of robots in the undergraduate curriculum: experience reports'. Proc. 32nd SIGCSE Technical Symp. on Computer Science Education, Charlotte, North Carolina, USA, February 2001, pp. 404–405

[Hirst03]: Hirst A.J., Johnson J., Petre M., Price B.A., Richards M.: 'What is the best programming environment/language for teaching robotics using Lego Mindstorms?', *Artif. Life Robot.*, 2003, **7**, (3), pp. 124–131

[Imberman07]: Imberman S.P., Klibaner R., Zelikovitz S.: 'Student feedback on robotics in CS1'. <http://www.cs.hmc.edu/roboteducation/papers2007/c1> (Accessed 11 March 2010)

[Jadud03]: Jadud M.C., Schleter J.: Little Languages and Little Robots. http://www.jadud.com/people/mcj/_les/2003-PPIG-II DBLP (Accessed 11 March 2010)

[Kurebay06]: Kurebayashi S., Kamada T., Kanemune S.: 'Learning computer programming with autonomous robots'. Informatics Education: The Bridge Between Using and Understanding Computers, 2006, 4226, Lecture Notes in Computer Science, pp. 138–149

[Ladd05]: Ladd B., Harcourt E.: 'Student competitions and bots in an introductory programming course', *J. Comput. Small Coll.*, 2005, **20**, (5), pp. 274–284

[Lauwers09]: Lauwers T., Nourbakhsh I., Hamner E.: 'CSbots: design and deployment of a robot designed for the CS1 classroom'. Proc. 40th SIGCSE Technical Symp. on Computer Science Education, Chattanooga, TN, USA, March 2009, pp. 428–432

[Lawhead02]: Lawhead P.B., Duncan M.E., Bland C.G., Goldweber M., Schep M., Barnes D.J., Hollingsworth R.G.: 'A road map for teaching introductory programming using LEGO Mindstorms robots'. Proc. ITiCSE-WGR '02: Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education, Aarhus, Denmark, June 2002, pp. 191–201

[Lemone96]: Lemone K.A., Ching W.: 'Easing into C++: experiences with RoBOTL', *SIGCSE Bull.*, 1996, **28**, (4), pp. 45–49

[McWhorter09]: McWhorter W.I., O'Connor B.C.: 'Do LEGO Mindstorms motivate students in CS1?' Proc.

SIGCSE ACM Technical Symp. on Computer Science Education, Chattanooga, TN, USA, March 2009, pp. 438–442

[Petre04]: Petre M., Price B.: ‘Using robotics to motivate “Back Door” learning’, *Educ. Inf. Technol.*, 2004, 9, (2), pp. 147–158

[Sartzemi03]: Sartzemi M., Xinogalos S., Dagdilelis V.: ‘An environment for teaching object-oriented programming: objectKarel’. Proc. 3rd IEEE Int. Conf. Advanced Learning Technologies, Athens, Greece, July 2003, pp. 342–343

[Sartzemi05]: Sartzemi M., Dagdilelis V., Kagani K.: ‘Teaching programming with robots: a case study on Greek Secondary education’. *Adv. Inf.*, 2005, 3746, Lecture Notes in Computer Science, pp. 502–512

[Sklar06]: Sklar E., Parsons S., Azhar M.Q.: ‘Robotics across the curriculum’. Technical Report, Department of Computer and Information Science, Brooklyn College, 2006

[Summet09]: Summet J., Kumar D., O’Hara K., Walker D., Ni L., Blank D., Balch T.: ‘Personalizing CS1 with robots’. Proc. 40th ACM Technical Symp. on Computer Science Education, Chattanooga, TN, USA, March 2009, pp. 433–437

[Vandelden08]: van Delden S., Zhong W.: ‘Effective integration of autonomous robots into an introductory computer science course: a case study’, *J. Comput. Small Coll.*, 2008, 23, (4), pp. 10–19

[Weiss08]: Weiss R., Overcast I.: ‘Finding your Bot-Mate: criteria for evaluating robot kits for use in undergraduate

computer science education’, *J. Comput. Small Coll.*, 2008, 24, (2), pp. 43–49

[Wellman09]: Wellman B.L., Anderson M., Vrbsky S.V.: ‘PREOP as a tool to increase student retention in CS’, *J. Comput. Small Coll.*, 2009, 25, (2), pp. 167–175

[Williams03]: Williams A.B.: ‘The qualitative impact of using LEGO MINDSTORMS robots to teach computer engineering’, *IEEE Trans. Educ.*, 2003, 46, (1), pp. 206

[Wu08]: Wu C., Tseng I., Huang H.: ‘Visualization of program behaviours: physical robots versus robot simulators’. Informatics Education – Supporting Computational Thinking, 2008, 5090, Lecture Notes in Computer Science, pp. 53–62

[Imberman05] Imberman S.P., Klibaner R.: ‘A robotics lab for CS1’, *J. Comput. Small Coll.*, 2005, 21, (2), pp. 131–137

[Wong01] Wong K.: ‘Teaching programming with LEGO RCX robots’. Proc. 18th Information Systems Education Conference, Cincinnati, Ohio, USA, November 2001

9 Appendix 2

See [Table 2](#).

10 Appendix 3

See [Table 3](#).

11 Appendix 4

See [Table 4](#).

Table 2 Electronic search results

	ACM	CiteSeer	EBSCO	ERIC	IEEE	ISI	Keele	AEI	BEI
‘amateur’	20	2	2	1	7	0	0	1	0
‘beginner’	42	5	1	0	4	1	0	1	1
‘first time’	575	3	1	8	2	77	0	0	0
‘introductory’	120	71	13	21	99	16	0	53	2
‘novice’	93	69	6	10	209	6	0	35	2
‘teaching’	34	111	17	21	75	97	0	18	6
‘learning’	26	346	21	23	58	258	0	16	3
abstract	379	385	94	0	12	336	0	9	0
title	4	25	7	1	4	10	0	0	0
total papers returned	1293	1017	162	85	470	801	0	133	14
included in SLR (before removal of duplicates)	20	8	1	0	3	8	0	0	0

Table 3 Data extracted from included studies

Study	Computer language	Nature of robot (Phy or Sim)	Char. of novices	Type of studies	Scale of studies (no. participants)	Are robots effective?
[Barnes02]	Java	Phy – LegoMind	University	–	–	unclassifiable
[Becker01]	Java	Sim	University	–	–	yes
[Bell08]	Robolab	Phy – LegoMind	High school	–	–	yes
[Borge04]	Java	Sim	University	interviews and discussions	unknown	yes
[Brauner10]	Scratch	Phy and Sim	High school	critique/questionnaire	31	yes
[Buck01]	Java	Sim	University	–	–	unclassifiable
[Cliburn06]	various	Phy – LegoMind	University	–	–	mixed
[Czejdo09]	Python	Phy – Scribbler	University	–	–	yes
[Enderle08]	C++	Sim	various	–	unknown	yes
[Fagin01]	Ada	Phy – LegoMind	University	–	–	yes
[Fagin03a]	Ada	Phy – LegoMind	University	comparative – learning with against learning without robots	938	no
[Fagin03b]	Ada	Phy and Sim	University	–	–	mixed
[Flowers02]	Java	Phy – LegoMind	University	–	–	yes
[Garrett05]	NQC	Phy and Sim	various	–	–	yes
[Goldweber01]	various	Phy – Several	University	–	–	mixed
[Hirst03]	various	Phy – LegoMind	various	–	–	yes
[Imberman07]	C++	Phy – LegoMind	University	critique/questionnaire	121	yes
[Jadud03]	scheme	Phy – LegoMind	High school	–	–	yes
[Kurebay06]	Dolittle	Phy – Custom	High school	critique/questionnaire	40	yes
[Ladd05]	Java	Sim	University	–	–	yes
[Lauwers09]	Java	Phy – iCreate	University	critique/questionnaire + analysis of grades + impact upon retention	72	yes
[Lawhead02]	Java	Phy – LegoMind	University	–	–	yes
[Lemone96]	C++	Sim	High school	unknown	15	yes
[Mcwhorter09]	C++	Phy – LegoMind	University	critique/questionnaire	78	mixed
[Petre04]	various	Phy – Several	various	interviews and discussions	unknown	yes
[Sartzemi03]	various	Sim	University	pilot lessons	20	yes
[Sartzemi05]	Robolab	Phy – LegoMind	High school	pilot lessons	unknown	mixed
[Sklar06]	various	Phy – Several	University	critique/questionnaire	unknown	yes
[Summet09]	Python	Phy – Scribbler	University	analysis of grades	259	yes
[Vandelden08]	Java	Phy – LegoMind	University	critique/questionnaire	unknown	yes
[Weiss08]	various	Phy – Several	University	–	–	unclassifiable
[Wellman09]	custom	Phy – iCreate	University	critique/questionnaire	40	yes
[Williams03]	C	Phy – LegoMind	University	critique/questionnaire	45	yes
[Wu08]	Java	Phy and Sim	High school	comparative – Phy. against Sim.	151	yes

Phy – physical robot

Sim – simulated robot

Table 4 Data extracted from articles found after performing the post-SLR search validation

Study	Computer language	Nature of robot (Phy or Sim)	Char. of novices	Type of studies	Scale of studies (no. participants)	Are robots effective?
[Imberman05]	C	Phy – Lego HandyBoard	University	–	unknown	yes
[Wong01]	various	Phy – LegoMind	University	critique/questionnaire + comparative – learning with v learning without robots	unknown	yes