Systematic Literature Review Protocol: Teaching Novices Programming Using Robots

Louis Major
School of Computing and Mathematics,
Keele University,
Staffordshire, UK
l.major@epsam.keele.ac.uk

December 2010

1 Change Record

Document History	
Version 1.0	First draft of Protocol
Version 1.1	Correction of spelling mistakes
Version 1.2	Minor revisions made after expert review of Version 1.1 Protocol
Version 1.3	Revisions made to Quality Assessment Forms
Version 1.4	Final Version of Protocol

2 Background

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. This study employs the Systematic Literature Review (SLR) methodology [6] in order to investigate the use of robots as tools to aid the process of teaching programming. The SLR is a trustworthy, rigorous and auditable tool [7] and is one that allows for existing evidence to be collected and summarised while identifying gaps in current research [6]. This work will form part of a larger project that is investigating innovative methods which can assist novice programmers in their studies. No past SLR has been found to examine the use of robotics in such a context. The Protocol that is presented will specify the plan that is to be followed in order to ensure the successful implementation of the SLR.

3 Research Questions

The objectives of the SLR are to:

- Undertake a systematic review of the literature related to the teaching of introductory programming using robots.
- Select a sub-set of studies to review in greater detail.
- Collect and analyse the evidence from these studies in order to assess the benefits and limitations of using robots to assist in the teaching of introductory programming concepts.
- Offer a clear picture of the current state of research in the field.

- To provide an appropriate framework/background in order that future research activities can be appropriately positioned.

The overall aim of the SLR is to determine:

"How effective has the use of robots been in the teaching of introductory programming?"

In order to achieve this objective, and to ensure the collection of all relevant data, several research questions have been created. These questions will help to ensure that the study will be comprehensive in its nature, whilst providing an in-depth analysis into the past use of robots to teach novices programming. The research questions are:

[**SLRRQ1**] What computer languages are being taught in introductory programming courses that make use of robots as teaching tools?

[**SLRRQ2**] Are the robots that are being used simulated or physical (real-life)?

[**SLRRQ3**] What are the characteristics (i.e. what is their age, level of education etc.) of the novices being taught?

[**SLRRQ4**] What types of studies are being performed by researchers that investigate the teaching of introductory programming concepts using robots?

[**SLRRQ5**] What is the scale of studies (e.g. number of participants) that are being performed by researchers?

[**SLRRQ6**] Do collected studies suggest that using robotics to teach introductory programming is effective?

4 Search Strategy

This section outlines the strategy that will be used to search for primary studies. The search strategy is intended to identify and collect all of the literature that complies with the inclusion and exclusion criteria that is detailed in Section 5. A mixed search strategy will be adopted and includes both automatic searches of electronic databases and manual searches of conference and journal proceedings. Past experience in performing a systematic mapping study was used to develop the search strategy that will be implemented. Note, year of publication will not be restricted as trial searches (Section 4.4) suggest that the devised search strategy will return a manageable number of results that will not need to be further reduced.

4.1 Resources To Be Searched

The following resources will be searched in order to identify and collect relevant material:

Electronic Databases
ACM Digital Library, CiteSeerX, EBSCOhost, IEEExplore, ISI Web of Science and Keele Universities Digital Library. The Australian Education Index, British Education Index and ERIC (Education Resources Information Centre) will also be searched to ensure that an adequate search of education based articles and journals has taken place.

Conference Proceedings
ECSS (European Computer Science Summit), ESOP (European Symposium on Programming), ICSE (International Conference on Software Engineering), ITiCSE (Conference on Innovation and Technology in Computer Science Education), SIGCSE (Technical Symposium on Computer Science Education) and SIGITE (Conference on Information Technology Education).

Journals
The Journal of Information Technology Education and the Oxford Computer Journal.

Other
Other sources of evidence will also be searched to ensure thoroughness. As the study progresses it is intended that reference lists from included studies will be manually examined to identify literature of interest that was not discovered during the initial search. Whilst a general search of the internet will not take place, due to returned results likely being un-citable, the internet will be used to find copies of specific papers.

4.2 Search Strategy

Full text searches will be run on each of the resources outlined. The following search strings have been devised and will be run on all the electronic resources discussed:

1. (robots OR robotics) AND
   ("amateur programming" OR "amateur programmer")
2. (robots OR robotics) AND
   ("beginner programming" OR "beginner programmer")
3. (robots OR robotics) AND
   ("first time programming" OR "first time programmer")
4. (robots OR robotics) AND
   ("introductory programming" OR "introductory programmer")
5. (robots OR robotics) AND
   ("novice programming" OR "novice programmer")
6. (robots OR robotics) AND "teaching programming"
7. (robots OR robotics) AND "learning programming"

A single Boolean search string can be devised by combining all of these search terms together:

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

After these searches have been run the following search string will then be 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 has been chosen in order to ensure that all relevant material will be identified and to make searches of the electronic resources more manageable.

4.3 Search Results and Documentation

As it is expected that a large number of references will be collected during the course of the SLR, the JabRef reference manager has been selected for the management of these. The reference of each paper that is deemed suitable for inclusion in the SLR will be entered into the reference manager along with the DOI (Digital Object Identifier), a copy of the abstract and a short note detailing why the paper has been accepted. Each reference will also be given a unique ID key.

The conduct of the search will be documented in order for the review to be transparent and repeatable. Any amendments to the search strategy will be detailed and justified. Search results that are returned will also be saved for later re-analysis (if required). If a publication is found

on more than one resource only the most recent will be accepted into the SLR. The full texts of accepted publications will be stored on two separate drives (with the file-name being changed so that it corresponds with the unique ID key that is stored by the reference manager).

4.4 Validation of the Search Strategy

During a systematic mapping study that was previously performed, several papers were identified as being relevant to the study that is presented here:

- Fagin, B. (2003) Measuring the Effectiveness of Robots in Teaching Computer Science. In *Proceedings of the 34th SIGCSE Technical Symposium on Science Education*, pp. 307–311. ACM.
- Flowers, T. R. and Gossett, K. A. (2002) Teaching Problem Solving, Computing, and Information Technology with Robots. *J. Comput. Small Coll.*, 17 (6), pp. 45–55.
- Petre, M. and Price, B. (2004) Using Robotics to Motivate 'Back Door' Learning. *Education and Information Technologies*, 9 (2), pp. 147–158.

Trial searches were run on the ACM Digital Library, IEEExplore and EBSCOhost electronic resources. As these papers were found during the testing process the inclusiveness of the search terms selected have been validated. Moreover, the quantity of results returned during the trial searches also provides an indicator as to how much effort will have to be exerted during the search for publications. This has been used to create a schedule that predicts how long the entire SLR process should take (see Section 9).

5 Selection Criteria

This section describes the inclusion and exclusion criteria that will be used to ensure that only relevant literature is accepted into the SLR.

5.1 Inclusion Criteria

- Publications will only be included that report on the use of robotics in teaching introductory programming to students who are studying either a specific Computing or IT-related course.
- Papers that involve an empirical study or have a 'lessons learned' (experience report) element will be included.
- Where several papers have reported the same study only the most recent paper will be included.
- Date of publication will not act as a barrier for inclusion.
- Grey literature (such as technical papers or government reports) will be accepted if relevant.

5.2 Exclusion Criteria

- Publications will be excluded if their main focus is 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.
- Papers that just propose an approach or describe the use of robots to teach introductory programming (with no 'lessons learnt' component) shall be excluded.
- Papers and reports will be excluded where only the abstract but not the full text is available.
- Publications will be excluded if they are not written in English.
- Letters, editorials and position papers will all be excluded.

5.3 Selection Process

The selection process will be performed by one researcher (the author) with a test/re-test approach being applied after an interval (of one month) to verify the selection process. In addition to this, a second researcher (a PhD supervisor) will independently select a random sample of the archived search results and compare these with the results of searches that they have undertaken. By adopting these techniques the consistency of the inclusion and exclusion criteria can be tested and verified.

The selection process is effectively divided into two phases:

1. Publications found during the initial search are assessed for their suitability based upon analysis of their title and abstract. Clearly irrelevant literature is excluded.
2. Publications selected during Phase 1 are then subject to a more thorough analysis (i.e. reading of the full text). This is done to ensure that the publication in question definitely contains information that is relevant to the study being performed, as well as data that can be extracted for later analysis.

5.4 Study Quality Assessment

Each publication in the final set will be assessed for its quality. The quality assessment procedure will occur at the same time as the extraction of relevant data and has the aim of ensuring that a particular studies findings will make a valuable contribution to the SLR. 11 criteria for quality assessment are discussed by Dybå and Dingsøyr in [8]. These criteria were used by Dybå and Dingsøyr in an SLR when there were a number of different study types. Use of the same criteria has been deemed appropriate for use during this SLR as it is envisaged that it will include examples of several different study types also. The 11 criteria that will be used to assess the quality of each publication are detailed below:

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 will be used to exclude non-research papers and those that do not clearly state the aims of their research. Literature that does not satisfy these two criteria will be excluded from the SLR. As such, these first two criteria represent the minimum quality threshold that will be observed during the review. The remaining nine criteria aim to determine the rigour and credibility of the research methods employed in a particular study 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, will be inserted into a spreadsheet and assigned a value of either 1 ('Yes') or 0 ('No'). Judgement will then be passed in relation to what papers report the most valuable findings and are most important to the study. In order to test the validity of the study quality assessment a PhD supervisor will be given a random set of seven papers and then be asked to assess these based on the criteria outlined. Any disagreements in the classification of quality will then be discussed and resolved by consulting a second PhD supervisor. See Appendix 1 for the complete study quality assessment form.

6 Data Extraction

This section describes how the information that is required from each publication will be obtained.

6.1 Data To Be Collected

All data will be extracted by a sole reviewer (the author). A second assessor (a PhD supervisor) will independently extract information from a random sample of publications. The results will then be compared and contrasted with one another. If significant anomalies are evident in the testing of this random sample, discussions will then take place amongst the assessors, in conjunction with a second PhD supervisor, to resolve the issue.

The following information will be extracted from accepted studies:

- Abstract and bibliographic reference along with the date of data extraction
- Why the study was accepted into the SLR
- Study type (e.g. journal paper, conference paper, technical report etc.)
- Study aims and objectives
- Setting of the study
- Methodology of study (e.g. observational, experience reports, comparative etc.)
- Information about baseline where appropriate (i.e. method against which robots are 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 novices being taught (e.g. age, level of education)
- Type of computer language being taught using robots (e.g. Java, C++, Other)
- Nature of the robot being used to teach a 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 robots as an intervention to teach introductory programming
- The study quality assessment (as discussed in 5.4)

To extract relevant information publications will be read in full. All extracted data will be stored within a spreadsheet. This will allow for easy reference when writing the report and will also ensure that all information pertinent to the study is stored in one location.

6.2 Data Synthesis

Until the results of the SLR have been aggregated the data synthesis strategy will not be finalised. It is envisaged, however, that data will be tabulated in a number of research question related tables. Following this, a narrative summary of the findings across the studies will be written. Data has already been extracted from the three papers identified in Section 4.4, during a trial run, in order to test the data synthesis process.

7 Study Limitations

An aspect of the study over which the author has little control is in relation to the electronic databases that will be utilised during the SLR. As the results are returned automatically by these electronic resources, little influence can be exerted over what is returned by them. This may result in no two searches being truly identical and have an impact upon the possibility of future studies being able to exactly replicate the research performed.

Several decisions that have been taken, during the creation of the Protocol that is presented, may also have an effect upon the results produced during the SLR. To elaborate, the inclusion, exclusion and study selection criteria, as well as the data extraction strategy, may all be influenced by different individuals interpretations (i.e. what one researcher included another may have excluded). The risk of such an event occurring, however, will be minimised by ensuring that relevant sampling of results is undertaken by an individual (a PhD supervisor) other that the main researcher (the author). A test/re-test method will also be implemented in order to supplement this sampling process.

8 Validation of the Protocol

Validation of the Protocol presented has been performed after an earlier version (1.1) was given to two expert reviewers (Professor Barbara Kitchenham and Professor Pearl Brereton both of Keele University). The changes that were suggested by these reviewers have been incorporated into the Protocol that is presented.

9 Schedule

Time	**Activity**	**Deliverables**
Week 1	Develop SLR Protocol, Give to Experts	n/a
Week 2	Revise SLR Protocol	Finish Protocol
Weeks 3-6	Search for Publications	List of Publications
Week 7	Test/Re-test Process	n/a
Week 8	Publication Selection	List of Selected Publications
Weeks 9-10	Data Extraction	Tables of Data
Week 11	Data Analysis	n/a
Weeks 12-14	Report Write-up	Finish Report

Appendix 1: Study Quality Assessment Form

Question & Considerations

Is this a research paper?
- does the paper involve an empirical study or is it a "lessons learned"/experience report based on expert opinion?

YES ☐ NO ☐

Is there a clear statement of the aims of the research?
- is there a rationale for why the study was undertaken?
- is there a clear statement of the studies primary outcome?
- is the main focus on the use of Robotics in teaching introductory programming concepts to Computing or IT students (e.g. those who are studying a specific Computing or IT course and are not being taught programming as part of another subject or the general curriculum?)

YES ☐ NO ☐

If the response to either of the above questions is NO the Quality Assessment procedure is not to be continued for that paper.

Is there an adequate description of the context in which the research was carried out?
- is the context in which the study was performed clear?

YES ☐ NO ☐

Was the research design appropriate to address the aims of the research?
- has the researcher justified the research design (e.g. have they discussed how they decided which methods to use)?

YES ☐ NO ☐

Was the recruitment strategy appropriate to the aims of the research?
- has the researcher explained how participants and cases were identified and selected?
- are these representative of an entire population?
- was the sample size sufficiently large?
- how might the recruitment strategy impact upon the results?

YES ☐ NO ☐

Was there a control group with which to compare treatments?
- how were the controls selected?
- were they representative of an entire population?
- was there something special about the controls?
- was the non-response high?
- could respondents be different in any way?

YES ☐ NO ☐

Was the data collected in a way that addressed the research issue?
- were all measures clearly defined?
- is it clear how data was collected?
- has the researcher justified the methods that were chosen?
- if the methods were modified during the study has a justification been given as to why?

YES ☐ NO ☐

Was the data analysis sufficiently rigorous?
- was there an in-depth description of the analysis process?
- has sufficient data been presented to support the findings?
- to what extent has contradictory data been taken into account?

YES ☐ NO ☐

Has the relationship between the researcher and participants been considered to an adequate degree?
- has the researcher examined their own role, potential bias and influence during formulation of research questions, sample recruitment, data collection and analysis?

YES ☐ NO ☐

Is there a clear statement of findings?
- are the findings explicit?
- has an adequate discussion of the evidence, both for and against the researchers arguments, taken place?
- are the limitations of the study discussed?
- are the conclusions justified from the results?

YES NO

Is the study of value for research or practice?
- does the researcher discuss the contribution the study makes to existing knowledge or understanding?
- does the research identify new areas where research is necessary?
- does the researcher discuss whether or how the findings can be transferred to other populations?

YES NO

Total Score

Note to the Reviewer:

If the response to a question is n/a (Not Applicable) please answer 'NO'.

A 'YES' constitutes a 1 and a 'NO' equates to a 0. Please consider this when calculating the overall quality score for a publication.

References

[1] Caitlin Kelleher and Randy Pausch. Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. *ACM Comput. Surv.*, 37(2):83–137, 2005.

[2] Kirsti Ala-Mutka. Problems in Learning and Teaching Programming - A literature study for developing visualizations in the Codewitz-Minerva project. 2004.

[3] Anthony Robins, Janet Rountree, and Nathan Rountree. Learning and Teaching Programming: A Review and Discussion. *Computer Science Education*, 13(2):137–172, June 2003.

[4] Sherry Turkle. *The Second Self - Computers and the Human Spirit (Twentieth Anniversary Edition)*. The MIT Press, 2005.

[5] Micaela Esteves, Ricardo Antunes, Benjamim Fonseca, Leonel Morgado, and Paulo Martins. Using Second Life in Programming's Communities of Practice. *Groupware: Design, Implementation, and Use: 14th International Workshop, CRIWG 2008, Omaha, NE, USA, September 14-18, 2008*, pages 99–106, 2008.

[6] Barbara Kitchenham and Stuart Charters. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 2007.

[7] Barbara A. Kitchenham. Procedures for undertaking systematic reviews. Technical report, Computer Science Department, Keele University, 2004. Available at: www.ebse.org.uk (Accessed 18 February 2010).

[8] Tore Dybå and Torgeir Dingsøyr. Empirical studies of agile software development: A systematic review. *Inf. Softw. Technol.*, 50(9-10):833–859, 2008.