Abstract

This paper reports on the thirteenth and final BRACElet workshop. In this paper we provide a brief retrospective review of the workshops and the findings that have resulted from this multi-institutional multinational investigation into the teaching and learning of novice programmers. Subsequently we report on the work undertaken during the final workshop and then discuss future avenues for research that have evolved as a result of the BRACElet project.

Keywords

novice programmers, SOLO, Bloom, assessment

1. Six years of BRACElet workshops 2004–2010

The workshop reported in this paper is the final workshop in a series of workshops. A brief outline of the outcomes of the previous workshops, partly quoting Clear, Philpott, and Robbins (2009), is presented below. A fuller discussion of the BRACElet projects findings can be found in Section 2.

December 2004: The inaugural workshop began at Auckland University of Technology (AUT) with a review of the findings of the Leeds working group which had recently been published (Lister et al., 2004). The participants decided to replicate and further the Leeds Group study by basing the question set design on an accepted educational model.
participants selected the revised Bloom's taxonomy (Anderson et al., 2001) and then devised a question set, including some of the original Leeds questions, which mapped to the "understand" cognitive process dimension. The questions were designed to test the students’ ability to reason about code and associated abstractions. After the workshop, as part of a pilot study, these questions were included in the final exam of the introductory programming (CS1) course in a number of participating institutions.

**July 2005:** During the second workshop held at the 18th Annual NACCQ conference in Tauranga, participants analysed, evaluated and published the results of the pilot study (Whalley et al., 2006). The results were also employed to further hone the research instrument. A second educational model, SOLO taxonomy (Biggs & Collis, 1982) was introduced to analyze and classify student responses to questions that required them to explain the purpose of a piece of code. By the end of the second workshop a toolkit had been developed that allowed the participants to undertake a fully-fledged study at their respective institutions. The results from this work were disseminated at ACE 2006 (Whalley et al., 2006) and further papers were authored over the next few months by subgroups.

**March 2006:** A third workshop took place at AUT during which the group developed a prototype ‘common framework’ (Lister, Whalley, & Clear, 2006) that allowed researchers to compare and contrast studies undertaken within BRACElet, but that also gave them the flexibility to tailor research to their particular interests.

**July 2006:** A fourth workshop was held at the 19th Annual NACCQ conference in Wellington. The common framework was reviewed and subsequently refined for the next phase of the project, an investigation into the code writing skills of novice programmers.

**July 2008:** A fifth workshop was held at the 20th Annual NACCQ conference in Nelson. It was at this workshop that a new type of question was introduced to the common framework pool, namely Parsons’ Problems (Parsons & Haden, 2006). This type of problem was devised by Parsons to help students acquire competence with the structural syntax of programming. Parsons’ problems, in essence, provide all the code required to solve a given problem but require the students to order (and possibly select then order) the lines of code to form a correct solution. A helpful paper on the use of these questions in assessment is provided by Denny, Luxton-Reilly and Simon (2008). The BRACElet group initially introduced the problem as an assessment question that we believed was assessing an intermediate stage between code writing and code reading. This assumption was proven incorrect in later work detailed in Section 3 when we investigated a possible hierarchy of programming skills using the common framework.

**December 2007:** A sixth workshop was held at AUT University, sponsored by ACM & SIGCSE. Contributed data was analysed. New research instruments were developed to allow for the evaluation of novice programmers’ program-writing skills and to enable comparisons to be made with their program reading skills. This toolkit was to be subsequently implemented at the participants’ institutions in Australasia, the data analysed and joint publication(s) produced.

**January 2008:** The seventh workshop was held at Wollongong at the ACE 2008 conference. This was the focal point for gaining Australian input related to the Carrick Institute Joint Associate Fellowship of Raymond Lister and Professor Jenny Edwards. An action research cycle was initiated at this meeting with the goal that this BRACElet iteration would: consolidate the “explain in plain English” question; generate new gender effects; examine any differences between international and local students (will have to be careful to differentiate between local “native speakers” and local ESL); relate answers to SOLO levels; examine differences between Undergraduate and Postgraduate; help develop an ideal exam paper.

**July 2008:** The eighth workshop was held at AUT University concurrent with the NACCQ Conference. The purpose of the workshop was to analyse assessment data from novice programmers, and thus further the inquiries into how novice programmers comprehend and write computer programs. Results from the workshop were suggested SOLO categories applicable to program writing tasks, and a modified SOLO classification scheme that combined reading and writing.

**September 2008:** The ninth workshop was held as a two day affair in connection with the ICER 2008 conference, in Sydney, and was supported by the Australian Learning and Teaching Council (formerly Carrick Institute). It was here that we started to talk about the value of replication studies to verify earlier findings. This workshop focused on the discussion of a relationship, if any, between reading/explaining code and writing code, and took its first step away from a folk-pedagogic debate, toward an evidence-based debate. It was at this juncture that the focus in BRACElet began to move beyond whether there was any relationship between reading/explaining code and writing code, to a discussion of the nature of the relationship – was it a causal relationship, or was it an indication of an underlying, yet to be identified, deeper skill?

**January 2009:** a half-day event, held at the end of the ACE2009 conference, in Wellington. This workshop was largely a reflection and discussion of progress to date. Because the data collection and analysis methods were well established it was at this point we decided to publish a paper to specify the next phase of data collection and analysis (Whalley & Lister, 2009). We also reviewed the draft of the group’s first replication study paper (Lister, Fidge & Teague, 2009). This paper replicated work by Lopez et al. (2008), using data collected from a different exam paper, conducted at a different educational institution. This study was later published (Lister, Fidge & Teague, 2009), but used a different statistical approach from Lopez et al. (2008). The results presented at the workshop were consistent with Lopez et al. (2008). Furthermore, the similarities and differences between this paper and the Lopez et al. paper led to a good discussion on what is essential, and what is not essential, to such a BRACElet study.

**June 2009:** A working group at ITiCSE in Paris provided the forum for the eleventh BRACElet workshop. The group replicated earlier analyses using a far broader pool of naturally occurring data, refined the SOLO taxonomy in code-explaining questions, extended the taxonomy to code-writing questions, furthered some earlier studies on students’ ‘doodling’ while answering exam questions, and explored a further theoretical basis for the work that until this point had been primarily empirical (Lister et al., 2009).

**January 2010:** The twelfth workshop was held in Brisbane as a half day event at the end of the ACE 2010 conference. BRACElet activities (now known as "threads") for the coming year were identified. The threads were as follows:

- A repetition of the classic Wiedenbeck study (1985) using a web based interface was instigated by Michael de Raadt (University of Southern Queensland) who also demonstrated a prototype of his web-based tool.
- Judy Sheard (Monash), Angela Carbone (Monash) Jacqueline Whalley (AUT, NZ), Mikko-Jussi Laakso (Yli-Ku, Finland) and Donald Chinn (University of Washington, USA) intend to carry out a survey and analysis of exam papers used to test novice programmers, looking at (for example) the types of exam questions used. They also plan to interview academics about their philosophy on teaching and assessment and how they set their exam papers.
- Mikko-Jussi Laakso (Finland, on sabbatical at Monash) demonstrated VILLE (Kaila, Rajala, Laakso & Salakoski,
2. Key Contributions: A retrospective

The key findings of the BRACElet project are founded on the analysis of naturally occurring data (see: Lister et al., 2010). First we present a list of the core findings related to novices learning to program.

1. Academics actively seek to abstract beyond the concrete code. Whereas the novices tended not to abstract, they "could not see the forest for the trees" (Lister et al., 2006).
2. A student's degree of mastery of code tracing tasks indicates their readiness to be able to reason about code. A student who has mastered the ability to trace code is also able to think relationally about code (Philpott, Robbins & Whalley, 2007).
3. The level of SOLO response to the 'explain in plain English' questions correlates positively with writing tasks (Sheard et al., 2008).
4. There is a correlation between performance on "explain in plain English" tasks and code writing tasks and between performance on code tracing tasks and code writing tasks. A stepwise regression, with performance on code writing as the dependent variable, was used to construct a path diagram. The diagram (figure 1, cited in Lopez et al., 2008) pointed to the possibility of a hierarchy of programming related tasks where knowledge of programming constructs forms the bottom of the hierarchy, with "explain in English", Parson's problems, and the tracing of iterative code forming one or more intermediate levels in the hierarchy (Lopez et al., 2008; Lister, Fidge & Teague, 2009).
5. There is a causal relationship, where a minimal level of skill at tracing is necessary for code writing, but that minimal skill at tracing is not sufficient by itself to enable code writing. It is the skills required for code explanation that, when combined with tracing skill, form a strong predictor of performance on code writing (Venables, Tan & Lister, 2009).

![Figure 1. A possible hierarchy of skills (Lopez et al., 2008).](image)

The BRACElet project has also contributed to research methodology in the field. They have published guidelines and toolkits for managing longitudinal multi-institutional multi-national projects (Whalley, Clear & Lister, 2007), in addition to providing methods that allow for the analysis of assessments and students code.

1. Guidelines for using Bloom's taxonomy to guide the design of programming assessments or programming problems (Thompson et al., 2008)
2. Guidelines for using SOLO to reliably classify student responses to questions that ask them to explain code (Clear et al., 2008)
3. Guidelines for applying SOLO to setting programming assessments and tasks (Thompson, 2010)

3. The last workshop

The purpose of the final workshop was to extend the work on theory initiated at the Paris ITiCSE Working group in 2009, and to launch an ongoing programme of research through the Software engineering Research Laboratory (SERL) at AUT into the novice to expert programmer continuum, the "N2Expert programmer continuum". This project will be titled the "NExpertise" project. The scope would extend from 'programming in the small' to 'programming in the large', and the work of professional programmers. Thus the work has the goal of contributing to broader and deeper understandings and impacting practice for both educators and software practitioners.

This two day workshop was held at Auckland University of Technology in September 2010, and was attended by twelve participants from New Zealand, Australia, Finland and Sweden.

The workshop began with a keynote given by Dr. Anna Eckerdal: "The delicate art of connecting theory and practice: A variation theory study in programming education". The second keynote of the workshop was given by Dr. Raymond Lister. He presented recent work, which will appear as two papers in the Australian Computer Education Conference in...
January 2011. His talk included a discussion of novice programmer reasoning and Concrete and other Neo-Piagetian theories (Lister, 2011) and work arising from BRACElet work on early relational reasoning in the novice programmer (Corney, Lister & Teague, 2011). The two keynotes were followed by a presentation on the preliminary data patterns and potential research designs for studying the development of expertise in novice programmers (BRACElet → NExpertise), arising from the data to be utilised in this workshop, given by Mike Lopez.

Following the presentations, five potential research questions the workshop leaders had prepared relating to previous BRACElet work were discussed. Subsequently the workshop participants formed three groups that focused on three of these questions.

1. Is there an ordinal relation between the TEXT, ACTION and MODEL aspects of the concepts object and class (Eckerdal & Thuné, 2005) in novice students’ learning?
2. Can we identify misconceptions and missing critical aspects from student examination responses?
3. Are there longitudinal trends in programming students’ learning? Is it possible to predict form exam results how students will perform later in their education?

Data provided, from courses spanning CS0, CS1, CS2 and C3, by the participants was then inspected in order to identify appropriate data and a preliminary analysis was undertaken to establish the feasibility of further each avenue of research.

**Group 1:** The group intended to investigate results from a previously performed phenomenographic study (Marton & Booth, 1997) on novice students’ understanding of the concept object and class (Eckerdal & Thuné, 2005). The results of the phenomenographic analysis gave an outcome space with three categories which illustrate qualitatively different understandings among the students in the study. Students demonstrated that they focused on one or more of the following critical aspects (Marton & Tsui, 2004) related to the two concepts. First, a text representation of the concepts, objects and classes are pieces of text with a certain structure. Next, an understanding that objects are active when a program is executed, and the actions are described in the methods of the class. Finally a few students also focused on the modelling aspect of the concepts. The phenomenographic study named these critical aspects TEXT, ACTION, and MODEL respectively.

The aim of this group’s investigation was to discover if it is possible to statistically refute the existence of a hierarchy in the development of novice students’ understanding with respect to the critical aspects TEXT, ACTION, and MODEL. The members in the group hypothesized that there exists an order between the three aspects in terms of how students learn them. The hierarchy starts with the TEXT aspect, then the ACTION and finally the MODEL aspect. The members further assumed that there is a common belief among educators that students learn to program according to this order. This implies e.g. that the ACTION aspect cannot be understood before the students understand the TEXT aspect. Such a belief can influence the way educators teach programming, and if the data refutes this belief, it is important information to the community of educators.

**Group 2:** There was a discussion among all participants of the workshop on research on misconception. Some researchers claim that, at least in Computer Science, there exist misconceptions that are so common that they are worth studying. Group 2 was interested in investigating students misconceptions of assignments (=) in Java, and how this relates to the representation of equality in mathematics. Are there any patterns in the errors related to assignment in students’ exams that can be understood in relation to the mathematical equality symbol and its conception? Closely related to assignment is state which is suggested by Shinners-Kennedy (2008) to be a threshold concept (Meyer & Land, 2005).

**Group 3:** The group was interested in performing a longitudinal study on individual students. The hypothesis was that students who perform well in introductory courses also perform well on subsequent courses. The group began to investigate the hypothesis by looking at students’ results in later courses, and correlating it to results for the same students’ in earlier courses. The research question is interesting in relation to common understandings among educators that students are "good" or "bad" and that students remain "good" or "bad" throughout their education.

Students who perform well in later courses have thus probably performed well in earlier courses. If the analyses showed the opposite, or no such pattern, the hypothesis is refuted.

### 3.1 Intermediate findings, issues reported and future work

**Group 1:** The group reported that they had tried several strategies on several data sets before one of the group members suggested a functioning strategy. The strategy the group used was to limit the analysis to try to refute a hierarchy between the first two aspects, TEXT and ACTION. The reason for this decision was mainly that the group found no exam question in the CS1 exams that reflected the MODEL aspect. If, however, the statistical analysis of the data refutes a hierarchy between the TEXT and the ACTION aspect, the hypothesis of a hierarchy between all three aspects is refuted.

The exam question that was chosen to illustrate the TEXT aspect was a Parsons’ puzzle question, and the question that illustrates the ACTION aspect was a question where a class Bus was given and the students had to predict return values of method calls after several other methods were called. The first question focused on code and the structure of the code, while the second question focused on the program execution.

A pilot analysis on 15 randomly selected exams was performed where all participants in the group were actively involved. A classification scheme was agreed upon. The results from the pilot study seemed to refute the hypothesis of a hierarchy between the two aspects TEXT and ACTION.

The group decided to split the analysis of the rest of the exam between the members in the group. The group will tentatively continue the communication in the group via email.

**Group 2:** The group examined interesting findings from an initial analysis of an end of semester exam for pre-degree students. A three part question involved nothing more than assignment statements, as follows:
Students had to state the value in each variable once the code had been executed.

For i, only half the students correctly gave the answer 4, 4. 30% gave r = 2, s = 4, seemingly ignoring the last line of code. For ii, 60% gave p = 8, q = 1 with only 12% giving the correct answer. For iii, 58% gave x = 7, y = 5, z = 3 with only 14% giving the correct answer. For the last two parts, most students seemed to take the values of the variables from the initial assignment and not track their changes.

In addition, in a CS1 exam, 6 of 10 students did not answer correctly on the question "What are the values of variables a and b after the following code has been executed?" The code is similar to ii) in the table 1: a = 2, b = 3, a = b, b = a. The wrong answer many students gave was: a = 3 and b = 2; they understood that the purpose was to swap values of two variables. The group will report interesting findings from an initial analysis of CS0 and CS1 exams.

The group will continue to analyse the problem with assignment in programming, and to point out the existing problems to CSE community. The group plans to design multiple experiments to investigate underlying issues behind this phenomenon. For example, we will conduct an experiment with two groups of students, a control group in which the participants answer the questions with pen and paper, and a second group which will utilize the same exercises in a digital learning environment (ViLLe, see Kaila et al., 2010). The goal is to find out what is the effect of the media (paper vs. digital) behind this phenomenon.

<table>
<thead>
<tr>
<th>i</th>
<th>int r = 2;</th>
<th>ii</th>
<th>int p = 1;</th>
<th>iii</th>
<th>int x = 5;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>int s = 4;</td>
<td></td>
<td>int q = 8;</td>
<td></td>
<td>int y = 3;</td>
</tr>
<tr>
<td>r</td>
<td>= s;</td>
<td>p</td>
<td>= q;</td>
<td>y</td>
<td>= x;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>q</td>
<td>= p;</td>
<td>z</td>
<td>= y;</td>
</tr>
</tbody>
</table>

The analysis emanated from a previous study on CS3 students’ assignment responses (Philpott, Clear & Whalley, 2009). This analysis categorised the students according to the SOLO taxonomy as previously interpreted (Clear, Philpott, Robbins & Simon, 2009/2010). Answers from certain assignment responses were categorised either as Unistructural, Multistructural, or Relational. The group identified students that had participated in the previous study, and looked for the same students’ exams from CS1 and CS2. The initial findings from the analysis on these students’ results for previous courses show a group of students who performed better on the CS3 assessment than on the earlier exams: the students were categorised as Relational on the SOLO scale on CS3, and lower on the previous exams. They also found a group of students who performed worse on the CS3 assessment, Unistructural on the SOLO scale on CS3, and lower on the previous exams. Preliminary findings show no patterns in the data. This can be due to the fact that many of the students who were analysed during the workshop were not ‘traditional’ students, but came from a wide variety of backgrounds. The group concluded that they need to analyse more data before they had any findings, and note that it is hard to longitudinally track students’ learning. If the lack of pattern still applies when a larger data set is analysed, this will refute the hypothesis that it is possible to predict from exam results how students will perform later in their education. The group also reported that there was no data available to follow the students’ performance in detail.

4. Where to from here?

As is typical with BRACElet workshops an agenda for follow-on work and publications has been charted (even for this "Final" workshop). All collaborations inevitably go through their phases and the parties to BRACElet have decided to formally bring the project to an end. This is timely with the 2010 publication of the results of the ALTC fellowship (Lister & Edwards, 2010), and reflects a diverging of interests apparent at the Jan 2010 workshop where new and separate threads have been spawned.

The formal cessation of the BRACElet project enables the Co-Principal Investigators to pursue their own lines of research, and begin some fresh enquiries. For the Auckland University of Technology Co-PI’s we see this direction spanning the full novice to expert continuum through the NExpertise project, in the hope that by better understanding the work of experienced professionals and intermediate programmers we may gain greater insights into how the fuller process of developing programming knowledge and skills develops and how these skills might best be taught and assessed. The differences between programming in the small and programming in the large are quite marked, so we will have many new questions, activities and contexts to consider. This will also raise new methodological challenges in determining how to study the more advanced programmer, in scoping research studies, in defining the artifacts required, and deciding how they may be collected and analysed.

There certainly seems to be scope for much fruitful work and we hope we can build another community of scholars working collegially on the NExpertise project, applying a model yet to emerge but similar to that we have promoted through the BRACElet studies.

In closing we wish to farewell the good ship BRACElet and thank all our colleagues who have sailed in her.

References


