



Bulletin of Applied Computing and Information Technology

Article B2:

05:01

June 2007

The Many Ways of the BRACElet Project

Jacqueline L. Whalley & Tony Clear

Auckland University of Technology, New Zealand

jacqueline.whalley@aut.ac.nz, tclear@aut.ac.nz

Raymond Lister

University of Technology, Sydney, Australia

raymond@it.uts.edu.au

Whalley, J., Clear, T. & Lister, R. (2007, Jun), The Many Ways of the BRACElet Project.

Bulletin of Applied Computing and Information Technology Vol. 5, Issue 1. ISSN 1176-

4120. Retrieved May 6, 2008 from

http://www.naccq.ac.nz/bacit/0501/2007Whalley_BRACELET_Ways.htm

ABSTRACT

This paper provides a retrospective snapshot of the first two years of a multi-institutional multi-national study (MIMN) in Computer Science Education called the BRACElet Project. This study has been inquiring into how novice programmers comprehend and write computer programs. The context for the study is outlined, together with details of how it has evolved and those who have participated. Some challenges encountered during the project are highlighted and pointers for the successful conduct of such a study are provided. The paper concludes by noting pitfalls to be avoided, some open research questions, and current plans for furthering the project.

Keywords

Multi-Institutional, Empirical, Computer Science Education Research, Collaboration, Novice programmers

1. INTRODUCTION

This paper provides a retrospective snapshot of the first two years of a multi-institutional multi-national study (MIMN) called BRACElet. We look at the evolution of the project, place the project in context and provide some useful guidelines as to how to manage such a project. Multi-institutional (MI) studies are becoming increasingly popular. The trend in computer science education (CSEd) research to adopt this genre of study can be explained by the pragmatic advantages of conducting research using the MIMN model. These advantages include statistical power, richness, hypothesis generation and improved methodology (Fincher et al., 2005). Indeed it was for these exact reasons that the BRACElet project started life as an MIMN study within Australasia.

2. HISTORY

BRACElet began as an investigation into the reading and comprehension skills of novice programmers. The study was motivated by the fact that current worldwide failure rates in introductory programming subjects are among the worst in universities. The difficulties that lead to failure appear to begin from the very first day of a student's introductory programming course. First year programming is the start of a rollercoaster journey for most students and while novices in every discipline make a similar journey we continue to wonder why the journey is so much more fraught with danger for programming students. We believe that this is partially due to the fact that we as educators are continually underestimating the difficulty of the tasks that we are asking students to undertake. Indeed, the fact that programming is more intensive in terms of cognitive load was illustrated in a study (Oliver et al., 2004) in which the cognitive difficulty level of six courses was analyzed using Bloom's taxonomy and a difficulty metric, called a *Bloom Rating* was computed. All the programming courses were reported to have a high overall *Bloom Rating* whereas other courses such as Networking were ranked significantly lower in cognitive difficulty.

Many papers in the current literature make it seem that our students' fragile grasp of programming and lack of problem solving skills is actually a new phenomenon and often this is linked to a paradigm or pedagogical change. In actual fact this problem is a recurring theme in the computer science education literature: students don't know how to read programs, they don't know how to design programs, they don't know how to problem-solve and they don't know how to write programs. The literature on learning to program is vast. A useful review of such literature can be found in a paper by Robins, Rountree, and Rountree (2003). Here we give a brief outline of the most relevant literature to date.

Twenty three years ago it was found that only 38% of computer programming students could write a simple program to calculate the average of a set of numbers (Soloway et al., 1983). Six years later Perkins and Martin (1989) reported that students had fragile knowledge of basic programming concepts and a "shortfall in elementary problem-solving strategies". An entire volume of papers, "Studying the Novice Programmer" also documented the difficulties of learning to program (Soloway and Spohrer, 1989). More recently a 2001 ITiCSE working group assessed the programming ability of an international student population from several universities (McCracken et al., 2001). The students were tested on a common set of program-writing problems and the majority of students performed more poorly than expected. It was not clear why the students struggled to write the required programs. In 2004, another ITiCSE working group (the "Leeds group") attempted to investigate some of the reasons why students find programming difficult (Lister et al., 2004). The working group set out to benchmark the program-reading skills of novice programmers. They found that many students could not answer program reading problems, "suggesting that such students have a fragile grasp of skills that are a pre-requisite for problem-solving".

3. THE WAY IT WAS MEANT TO BE

In New Zealand the BRACElet project arose as a logical follow on from an NACCQ Conference keynote (Lister, 2004), with a programme aimed at developing a supportive research community through which to further investigate the open issues highlighted by the Leeds study. Inspired by the Bootstrapping and BRACE models, it has also had from the outset an explicit CSEd Research development goal for those involved.

The BRACElet study was intended to extend the work of the Leeds group. The intention was to add new research instruments, to the existing Leeds toolkit, that would assess program-writing skills and enable comparisons to be made between students' program

comprehension (reading) skills and program writing skills. The project was expected to provide new insight into the little understood area of the relationship between comprehension and writing of programs by novice programmers.

As interesting as the results were from the Leeds group, that project did not have a theoretical underpinning. The choice of the reading problems was not based upon a theoretical model. The analysis of the data was not driven by any theory or model of how students should solve or actually address problems.

As a first phase, our intention was to replicate the Leeds group work, (which investigated the skill of students in reading and understanding computer programs), to give a baseline from which to subsequently investigate skills in the writing of programs. But our question set would be developed within the context of a theoretical framework to inform problem choice, problem-set design and data analysis. Common problem-sets for program reading and writing would be used by all participants and would be administered under comparable conditions to novice programming students. These common problems sets would eventually provide a pool from which we could draw for future studies. Additionally the experimental toolkit was to be designed so that we could collect opportunistic data such as script annotations and student interviews. Interviews had been used by the Leeds group to add richness to the data collected by capturing the students "thoughts out loud" with the intent of gaining a greater insight into a student's cognitive processes. At an early stage in the project a research toolkit was to be developed that would provide guidance for participants. The toolkit development would be guided by the excellent experiment kits (Fincher, 2002) provided for the participants of the Bootstrapping (Petre et al., 2003), Scaffolding (Fincher et al., 2004) and BRACE (De Raadt et al., 2005) projects.

We anticipated that subsequent cycles of research would result in the extension of the study or further replications of the study across other countries and institutions.

4. THE WAY WE WERE

We began work during a two day workshop in early December 2004 (Table 1). The BRACElet participants (see Table 2 for a list of participants) looked at the questions used by the Leeds group as well as those developed by Wiedenbeck et al. (1999).

This first workshop was focused primarily on developing the research instrument and toolkit. We decided to maintain the multiple-choice question (MCQ) format used by the Leeds group in our study in order to ensure that we had no variation in our data analysis between markers and institutions. The results from the MCQs would provide us with a student performance benchmark that we could use to look at other question types.

Table 1. Project timeline to date, the numbers in [] indicate a workshop or subgroup meeting id.

2004	<i>July</i>	Initial funding secured from AUT University
	<i>September</i>	Recruitment Preparation
	<i>December</i>	First Workshop [1] - Developed question set - Developed framework - Established working group protocols First pilot ethics application made by AUT team with Ray Lister's UTS application for inspiration
2005	<i>January</i>	Refined problem set finalised and localised by individual institutions. AUT receives ethics approval

	<i>January</i>	Other institutions begin ethics process.
	<i>May/June</i>	First data collection phase [AUT, Unitec, BoPPoly]
	<i>July</i>	Project briefly profiled in SIGCSE Bulletin Column (Clear., 2005)
	<i>July</i>	Data analysis commenced
		Second workshop [2] in collaboration with NACCQ at BoPPoly Tauranga
		- Reclassified the problem set
		- SOLO taxonomy introduced
		- Potential publications identified
		- Initial data reviewed
	<i>August</i>	Paper writing commenced
	<i>September</i>	Second data collection phase using common problem set
	<i>October</i>	Project situated as an MIMN study in ICER paper (Fincher et al., 2005)
2006	<i>January</i>	First paper reporting results presented at ACE 2006
		Subgroup meeting in Hobart, Australia during ACE 2006 [2a]
		new international members recruited
		common framework proposed
	<i>March</i>	Third workshop [3] – AUT University
		- Further pool of question types
		- Develop code writing questions
		- Revisit the existing framework for use as a common framework.
	<i>April</i>	Beth Simon prepares and runs the first exam script using the common framework at UCSD,
	<i>May</i>	Further exam scripts were prepared and run by other participants
	<i>June</i>	Paper presented at ITiCSE in Bologna, Italy.
	<i>July</i>	Two papers presented at the NACCQ conference in Wellington, New Zealand.
		Paper wins CITRUS award for collaborative research at NACCQ
		Fourth workshop [4] in collaboration with NACCQ at Whitereia Polytechnic, Wellington
	<i>August</i>	University of Auckland ethics application submitted

Table 2. Project participants (project leaders are identified by an *)

Institution	Participant	Year joined	Workshops attended	Co-Author
Auckland University of Technology (AUT) (AUT is the project's lead site)	Tony Clear*	2004	1,2,2a,3 & 4	Y
	Jacqueline Whalley*	2004	1,2,2a,3 & 4	Y
	Phil Robbins	2005	3 & 4	Y
	Gordon Stegink	2004	1	
	Gordon Grimsey	2004	3	
	Graham Bidois	2005	2 & 4	
	Anne Philpott	2004	3	
University of Turku, Finland	Linda Grandell	2006	2a	
	Mia Peltomäki	2006	2a	
Bay of Plenty Polytechnic	Ajith Kumar	2004	1,2,2a,3 & 4	Y
Eastern Institute of Technology	Cathy Saenger	2004	1	
Massey University (Auckland)	Heath James	2006	2a, 3	
Massey University (Wellington)	Errol Thompson	2004	1,2,2a,3 & 4	Y
Manukau Institute of Technology	Bob Gibbons	2004	1 & 2	
	Mike Lopez	2006		

National University of Ireland, Maynooth, Co. Kildare, Ireland	Des Traynor	2006	2a	
University of Southern Queensland, Australia	Michael de Raadt	2006	2a	
Tairawhiti Polytechnic	Minjie Hu	2004	1	
Otago Polytechnic	Joy Gasson	2004	1, 2 & 4	
	Dale Parsons	2006	2a	
Waiariki Institute of Technology	Don Kannangara	2006		
Waikato Institute of Technology	Hamiora Te Momo	2004	1	
Unitec	Christine Prasad	2004	1, 2 & 3	Y
University of Auckland	John Hamer	2004	1 & 3	
	Andrew Luxton-Reilly	2004	1 & 3	
University of California, San Diego, USA	Beth Simon	2005		Y
University of Technology, Sydney, Australia	Raymond Lister*	2004	1, 2, 2a & 3	Y

Prior to developing the new questions we decided that we would need to adopt some sort of cognitive framework in order to be able to measure how difficult each problem was. As a group we agreed on the revised version of Bloom's taxonomy (Anderson et al., 2001) to formalise the design of the instrument. The draft instrument we developed in this inaugural workshop incorporated a classification for each of the MCQs. Because we were aiming to measure student comprehension of code read, we developed problems, in sub groups and individually, that we believed would lie within the *Understand* class of the revised Bloom's taxonomy.

We adopted two of the more successful problems from the Leeds working group as a starting point. From there we developed, during the course of the workshop, a further set of MCQs. Our problem set was further extended to include a set of questions that were more open, subjective questions designed to test a higher level of the Bloom taxonomy. An example of a more open question is a question that asks a student to explain the purpose of a code snippet. This type of question was designed to assess higher cognitive levels than the MCQs. We then verified and confirmed, as a group, the Bloom category assignment for each question.

5. THE WAY IT HAS EVOLVED

Since the inaugural workshop, workshops have become a biannual event and provide participants with a chance to work together as a group. Work continues in the periods between with subgroup meetings at international conferences about twice a year when numbers permit.

Each workshop to date has involved developing a new piece of the experiment toolkit. In the inaugural workshop we developed a set of questions that were subsequently refined and a common experiment set finalized for the pilot data collection phase. The second workshop involved looking at the pilot data and some initial analyses. A common data repository was set-up as well as a specification for data formats. A schema for anonymous script scanning and data source identification was established. Additionally the SOLO taxonomy was introduced for the first time as a way of classifying student responses to short answer questions. In the third workshop we developed alternative question types for the problem pool and revisited the Bloom component of the framework. It was at this workshop that the

notion of a common framework rather than a common instrument was introduced.

5.1 From Common Instrument to Common Framework

As the project grew and more participants came on board, particularly internationally, we discovered that it became difficult to achieve consensus on a single experiment design, among so many people. Many factors contribute to the difficulty including different programming languages, pedagogy, ethics' processes and teaching and research environments.

The solution (Lister, Whalley and Clear, 2006) was a common framework for a set of concurrent experiments rather than a single experiment over multiple institutions. This common framework would be agreed upon and should allow participants to compare and contrast between experiments. Moreover such a framework should also give people the flexibility to tailor an experiment to their particular interests. The framework would be specifically for use in studies of programmers and in particular, student programmers. The proposed framework was agreed on in theory by a sub-group that met during the ACE 2006 conference. This framework has three essential components:

- A reading and code-tracing component with questions that are similar to the MCQs (1 to 9 in the original BRACElet problem set). The aim of this component is to rate the subjects ability to read and understand code, with categorisation by the Bloom taxonomy as a level of cognitive difficulty indicator. The data analysis from the component gives us a performance measure or score for each student that gives a measure of a student's ability to comprehend some aspect of code.
- A SOLO (Structure of the Observed Learning Outcome, Biggs, & Collis, 1982) component that aims to classify a student's response in terms the level of abstraction. For example, a question for this component may ask the students to explain the purpose of a piece of code or to recognise the similarities in code segments and apply a classification strategy.
- A writing component that allows us to rate a student's ability to write code, these writing questions should be put into the context of the framework by identifying a similar reading task that can be compared with the writing task and by classifying the question using the Bloom taxonomy.

An optional component for recall was also proposed. For this component we proposed that the participants not only run their experiments with their students but also with their colleagues. The recall component has its foundation in a classic psychology experiment. It is used to assess people's high-level understanding of something by testing the speed with which they memorize something. Such a study was made of chess players (Chase, & Simon, 1973) who were asked to memorize board positions of several chess pieces. Novices tended to remember the position of each piece in isolation, whereas experts organized the information at a more abstract level, the attacking and defensive combinations. When recalling board positions that arise naturally in chess games, experts outperformed novices, but with unnatural arrangements of pieces, the performance of the experts decreased because the abstract patterns that the experts typically used were not present in the unnatural arrangements. For programming, the differences between novices and experts have also been studied extensively, and tend to confirm the findings from other disciplines. Expert programmers form abstract representations based upon the purpose of the code whereas novices form concrete representations based on how the code functions. In a study of programming that reflected the earlier chess studies, Adelson (1984) showed that, when given typical tasks on well-written code experts outperformed novices, but when faced with unnatural tasks, novices sometimes outperformed the experts.

5.2 The Trouble with Bloom

Designing a set that tested the full range of cognitive processes within the *Understand* cognitive domain of the Bloom's Revised Taxonomy (Anderson et al., 2001) proved more difficult than anticipated. Indeed after the inaugural workshop we thought we had our classifications correct. However, while the revised taxonomy was a fertile source of ideas for generating questions, once a question was written it was sometimes difficult to formally place it within the revised taxonomy. The examples given by the taxonomy's authors are not easy to translate into the programming domain. In many cases the categories within the knowledge domain, did not readily fit with concepts and tasks required in computer programming. It was difficult to match the cognitive tasks undertaken for each question with Bloom's cognitive processes. This resulted in the working group initially underestimating the cognitive difficulty with respect to the revised Bloom's hierarchy of classifications.

At the group's second meeting it was realised that the group had initially categorised the questions according to what we would do rather than what the students would do when attempting the questions. A review of our categorisation was undertaken over three sessions by a consensus of six members of the working group. This recategorisation assumed that the Bloom categories represented a normative model of good practice carried out by students. During a presentation at the ACE conference in 2006 the conference delegates were asked to assign a subcategory within the revised Bloom taxonomy to a given MCQ. There was no consensus regarding the cognitive level of the question. This exercise demonstrated the difficulties in adopting Bloom for the framework.

Clearly we need to find a shared understanding of Bloom's Taxonomy so that we can compare the cognitive difficulty of questions across different instantiations of the framework.

Some attempts have been made to relate Bloom to specific computer programming tasks. The IEEE Swebok® Guide to Software Engineering (Abran et al., 2004) employed the original Bloom's taxonomy (Bloom et al, 1956). Schneider and Gladkikh (2006) attempted to develop a revised Bloom's taxonomy for use in planning programming assessments. They adopted the interpretation of the cognitive thinking model from Anderson,, & Krathwohl (2001), but found that the original terminology of Bloom's taxonomy was more representative of the tasks performed in IT and science disciplines. When looking at the categorizations proposed by Schneider and Gladkikh we find that our problem set falls into and covers all the subcategories within the 'analysis' level rather than the 'understand' level as we had first estimated. This gives support to our claim that our first classification underestimated the cognitive difficulty of the questions. Schneider and Gladkikh's "Modified Taxonomy" version of the Bloom taxonomy now appears to provide us with a good starting point from which to reach a shared understanding and taxonomy that can be reliably applied across different experiments, to enable valid comparisons of results.

Nonetheless we have found evidence that the Bloom classification or cognitive difficulty level of a question has a correlation with student performance. The higher the Bloom level the less likely a student is to arrive at the correct answer (Whalley et al., 2006). This was an encouraging finding as it suggests that educators can apply a "level of difficulty" yardstick with some granularity when setting programming MCQs.

While applying Bloom's taxonomy has proved challenging, the SOLO component of our framework works with ease. We have been able to reliably validate each other's experiments by blind SOLO classification of student responses (agreement is seen in over 95% of cases) and easily reach a shared understanding of the SOLO taxonomy (Whalley et al, 2006, Lister et al., 2006, Thompson et. al 2006). Because SOLO adapts so naturally to programming tasks, we have more recently been focused on this aspect of the framework, however focus

is now shifting back to a Bloom's taxonomy that allows the reliable classification of questions that are assessing program reading and those that are assessing code writing.

6. THE WAY IT WORKS

6.1 Key Elements of MIMN Projects

The table below indicates the key elements of the project as originally categorised in Fincher et al. (2005).

Table 3. Project Characteristics (excerpted from Fincher et al., 2005)

	BRACElet
Researcher Recruitment	Sectorally based – (NACCQ) - NZ Tertiary computing educators & colleagues from Leeds WG
Introduction of study to research team	Presented 2004, refined 2005 during pilot studies
Data	Multiple (categorised MCQ's, short answer questions and "doodles")
Analysis	Statistical and qualitative
Follow-up work	Is itself an adaptation/extension of Leeds WG
Model strengths & weaknesses	Builds upon prior work and shared expertise with key common participants. Enables mixture of novice and experienced researchers to work together. Costs for PI's in coordinating, hope to share data analysis load and writing.

Since then some new aspects have been added: in particular the mode of communication and the degree of exclusiveness of the project are distinct aspects of the project's operation. Given the distributed nature of the project and the amount of collaboration and co-authorship involved, the team has used not only email but also Skype™ for synchronous voice and text conferencing sessions, through which to jointly author and consolidate a set of study findings. For instance, the paper by Lister, Simon et al. (2006) involved some lengthy Skype™ sessions with Sydney, Auckland, Wellington and San Diego resident authors jointly discussing the contents and findings, to bring their ITICSE 2006 paper to fruition.

As Table 2 above indicates, the project is also an open project, which invites new members to join from time to time, and sees some members come and go. Yet there remains a core leadership and group of active members who collaborate to propel the project to achieve meaningful outcomes. Having said that, the group to date has operated on a consensus basis with team members taking initiatives as possibilities within their institutions present and as their confidence grows.

Two of the goals of MIMN studies in CSEd research typically relate to 1) informing the teaching-research nexus; and 2) providing community and support for often isolated computing educators. These goals are fundamental to this form of research.

6.2 It Works Better When There Are Agreed Rules of Engagement

What really drives BRACElet is a set of "Rules of Engagement" that cover the brainstorming or ideas phase and the post-brainstorming phase. The basic principle of these rules is that ideas are the easy part. They recognise that the hard graft is done during the ethics clearance, drafting the actual final instruments, collecting the data, and doing the analysis.

During the brainstorming phase the entire group across national boundaries, has a frank exchange of ideas on the many possibilities for implementing each framework component. If person A offers up an idea that is eventually used by person B then person A is *not* by default a co-author (also cf. Appendix A for suggested guidelines).

For the post-brainstorming phase sub-teams of collaborators/co-authors nucleate around specific, common instantiations of the components, for which all those "nucleus" team members collect data. It had been suggested that geographical locale determine these subgroups. For example New Zealanders (or people from any single country) might very naturally form a single nucleus. Such team nuclei may also, at **their** discretion, invite specific people from the broader framework to join their team (as coauthors) to contribute to data analysis and writing. However, in practice, the "nuclei" teams have evolved as international teams without borders and tend to form due to strong working relationships and common areas of interest. Due to the borderless nature of the "nuclei" teams a requirement of participation in the project is that each researcher working in the framework must set-up and become familiar with Skype™. In general, we do not discuss substantive issues via email because it is just too difficult and too slow.

6.3 It Works Better When There Is a Publication Protocol

It is important to develop and agree on a protocol for individual writing and external collaboration. The guidelines below outline the policies for joint authorship written (Lister, Whalley, & Clear, 2006) at the ACE2006 subgroup meeting, at which the common framework was proposed.

The authoring policies put in place for the common framework based experiments were required as the common framework created a new set of circumstances as regards experiment design and administration. It was suggested that conference papers be authored in arbitrary subgroups but that people should still strive to have results from more than one institution where appropriate. If a proposed publication uses data from more than one institution then, in the first instance, authorship is opened up to the entire group on the understanding that each author has the time to commit to meeting their obligations within the authorship team as outlined in Appendix A. If a paper is being prepared for conferences that group members are attending there is a managerial level of co-ordination of papers to ensure that they do not detract from each other or report similar work.

Post conference, the group discusses whether authors of the conference papers will subsequently form a single group to write a joint journal review paper. This is a consensus decision between the conference paper authors. One positive aspect to such authorship policies is that it gives appropriate recognition to multiple participants, by providing all with a chance to be first author. Our practice to date has been consistent with the guidelines and represents the current project consensus. In addition to these guidelines we adopted the authorship guidelines, set out in appendix A below, which were adapted from the British Sociological Association by Sally Fincher in mid 2005.

In the following section we give a brief retrospective of the advantages of MIMN and the key elements we have found to contribute to success. Since the publication protocols were first agreed we have added two new components:

- Until a joint paper is accepted for publication and the next iteration is started, BRACElet collaborators should not publish independently of the whole group. After that they are free to do so. However people should inform the rest of the group of their intention to publish and where practical invite BRACElet people to join them. Furthermore external collaborators are not free to subsequently work with the data on

their own. Otherwise ethics clearance issues arise for existing participants.

- Any individual publications should acknowledge the work of the BRACElet group.

6.4 It Works Better If Ethical Issues Are Addressed

A critical consideration is the need to first navigate the human subjects ethics process in each institution. This may take some time to complete and can delay studies for a semester or more. The participants identified some key points for the ethical application approval process:

1. Codes must be in place such that institutes and study subjects cannot be identified and that the data collected may not be used for inter-institutional ranking purposes
2. Appropriate security measures must be in place for storage of data and code-identification information
3. There must be a purpose for *all* the data collected and you must have an idea of what you are going to do with all of the data collected
4. Ethnic background or English as second language questions should be avoided due to ethics considerations
5. Subject approval forms should be prepared and submitted with any ethics application.
6. Subject coercion must be avoided, especially if you are the lecturer on a course the subjects are enrolled on.
7. Performance benchmarking should be included (even though this will be addressed at a later stage in the project)
8. A minimum subject recruitment number should be determined (e.g.: at least 20, determined by the number you need for statistical analysis).
9. The ability to offer incentives must be checked with your own ethics committee
10. Treaty of Waitangi, within New Zealand, one potential ethical issue in that the instrument and related package may have to be available in Maori as well as English.

Some institutions have categories of exemption for certain forms of research, which are less onerous than a full application. It pays to seek advice whether your proposed study may fit within such a category. There are paper level approvals in some institutions, and one application currently in train at the University of Auckland is aiming for a general approval to anonymously analyse student examination data, gathered in the normal course of teaching activities. We agree with the views of Zeni (1998) who asserts that most educational action research should be exempt from formal ethical review processes, and urges "academic institutions to support reflective teaching and to minimise the bureaucratic hurdles that discourage research by teachers to improve their own practice". However, that is not the environment within which we teach and conduct research, so we need to carefully ensure that we meet the necessary requirements.

One advantage of this MIMN study has been the ability to aid those researchers who have either not been able to finalise the ethics approval process, or who have had difficulty enlisting the support of colleagues and students. The latter requires a proactive approach, if the researcher is not the actual classroom teacher, since motivating the work and gaining the buy-in of colleagues is critical. The student culture in some institutions may also be quite resistant to additional work that is not summatively assessed, so studies requiring students to volunteer for additional work may prove problematic.

The use of embedded instruments occurring as a natural part of course assessment is likely to prove a more effective data gathering strategy. To retain the required voluntarism and informed consent provisions of an ethical research process, a design in which students have the option to consent or otherwise to further use of the data, can be an effective model. With the BRACELet project to date, we have been able to share the data analysis based

upon data from those first institutions that have been able to collect results. Thus team members who would otherwise have had no data, or have seen their study fail to collect sufficient quality data, have been able to make significant contributions to the work; in framework and instrument development, data analysis, and co-authoring of academic reports and papers.

6.5 It Works Better When Beliefs about the Teaching/Research Nexus Are Shared

Along with the managerial policies and practices and the formalised research instrument development, data gathering and analysis processes, as outlined in this paper, that lead to the success of a MIMN study, we must not forget that the real key to successful collaborations lies in the people involved in the project. The attitude of participants in CS Ed Research MIMN studies matters too. It is important that participants in the research believe in a strong teaching-research nexus and see this work as one way in which to improve the quality of one's own teaching, by evidence-based and research-informed model of practice. To be more explicit about our own stance in this paper, it may be useful to refer to the study of educators' perceptions about research and teaching by Robertson, & Bond (2001). They identified five distinct conceptions of the interrelationship between teaching and research, depicted in Table 4 below:

Table 4. Conceptions of the interrelationship between teaching and research

1	Research and teaching are mutually incompatible
2	Little or no connection exists between research and teaching at undergraduate level
3	Teaching is a means of transmitting new research knowledge
4	Teachers model and encourage a research/critical inquiry approach to learning
5	Teaching and research share a symbiotic relationship in a learning community

The authors believe strongly in the latter three of these conceptions.

6.6 It Works Better When There Are Protocols for Data Collection and Management

To enable meaningful data analysis and comparisons to be made between data from different sources and institutions, the importance of clear data management, common formats databases etc. cannot be overstated. For instance, in the second workshop we created a spreadsheet to serve as a common cross-institutional template for data collection, which included the necessary fields and their definitions.

While trying to operate as a joint project with common instruments, it is still necessary that each institution implement the instrument as best fits within their ethical, departmental and subject based constraints. Since the process adopted at each institution is relevant to analysis and interpretation of the data collated, it is also necessary to capture information regarding the use of the instrument to allow appropriate analysis. Additionally each institution should decide whether the students used in the study participate on a voluntary basis, or whether the instrument will be embedded into the course assessment.

Our own research typically emphasises a strong practice linkage and the practices we have adopted for data collection and analysis have been described in sections 3, 4 and 5 above.

6.7 It Works Better When Participants Can Meet

Unless an MIMN project can gain access to some level of seed funding to support regular meetings of collaborating parties, it will struggle to gain momentum. A certain amount can be done in conjunction with common conference attendances, especially where those conferences have accompanying workshops or other group time allocations. Once underway the group can also work virtually to good effect. However there is a need to meet regularly as a group, or even as subgroups, to discuss development, analyse findings, develop new instruments and plan the next stages of the research to maintain the continued health of the collaboration.

A gratifying outcome of the work has been the breaking down of pedagogical and research isolation that this collaborative project has engendered. It has resulted in a strong community of practice (Wenger, 1998), with much sharing of assessment items, advice and feedback on proposals by educators working in their own institutions. It has built a trusted and supportive set of relationships which mean that each educator now has a mentoring and support network upon which to call for resources advice and guidance.

MIMN projects need clearly defined research goals around which to craft joint frameworks and instruments. The ability to reflect, relate findings to the literature and adapt as the project develops is vital. Patience is a key quality as is perseverance. The project members feel that we are still coming to grips with the cognition of program reading, and are only just beginning to move into understanding the writing of code. Our ability to engage in a series of face-to-face meetings has helped to maintain a sense of cooperation and progress.

Furthermore our observations of participants in the project are that all have been challenged to consider deeply their own teaching and assessment practices, and have been inspired to innovate consciously and deliberately based upon insights from this work. The enthusiasm for good teaching and learning practices that has been generated by this work is truly gratifying.

7. THE WAY FORWARD

It has taken us longer to get our heads around program reading and comprehension than initially anticipated and this to date has been the primary focus of the BRACElet group.

At ACE2006, it was suggested that we might use more than one writing task, say one easy, one middle, one hard that all students attempt in order, from easy to hard, to avoid floor and ceiling effects.

Like McCracken (2001), for reliability of marking we will need to develop some sort of scoring rubric (or one rubric for each instantiation of the framework).

We are still focusing on small synthetic pieces of code (decontextualised) vs. larger, richer bodies of code at higher levels of granularity, but this richer code may be addressed at a later stage of the work. We are also still wrestling with the design of Object Oriented versus procedural instruments. The impact of these programming paradigm distinctions needs further work. What are the implications for assessment items that measure performance fairly for each paradigm, and is it possible to define combined items that measure both? The contested space around such issues (cf. Lister, Berglund, Clear, et al., 2006) may render this task difficult.

The primary Leeds Group paper (Lister et al., 2004) ends with a proposition for a follow-on

experiment – that students be given both reading tasks and writing tasks, to see if student performance on reading and writing correlate. We are currently considering an experiment that implements the following refinement to that Leeds Group proposition. The reading tasks should group students on whether they tend to respond (according to the SOLO taxonomy) multistructurally or relationally. The reading performance of each SOLO group should be correlated with the writing tasks. We suspect that the correlation for the students who tend to respond relationally will be higher than for the students who tend to respond in a multistructural way.

The use of the common framework has still to be validated, and its effectiveness in both sustaining and growing research collaborations and managing both joint and multiple publication processes has still to be proven. Yet based upon progress to date, the enthusiastic mood within the project team and the degree of mutual support we have generated, the authors are optimistic of success.

8. ACKNOWLEDGEMENTS

The BRACElet project gratefully acknowledges the financial support of AUT University through grants from the School of Computer & Information Sciences, and the Faculty of Design and Creative Technologies. The support from both NACCQ in New Zealand and the SIGCSE community internationally has also been appreciated.

We would like to thank the members of the BRACElet team without whose insights, hard work and active support, there would have been no progress to report. The role of those colleagues, who have helped with data collection, shared course materials and ideas, contributed to discussions at workshops and to instrument design, and who have not yet been recognized through publication, must also be acknowledged.

REFERENCES

- Abran, A., Moore, J., Bourque, P., DuPuis, R., & Tripp, L. (2004). Guide to the Software Engineering Body of Knowledge - 2004 Version -SWEBOK®. Los Alamitos, California: IEEE-CS - Professional Practices Committee.
- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 483-495.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (eds). (2001). *A taxonomy for learning and teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman Inc.
- Biggs , J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.
- Bloom, B., Englehart, M., Furst, E., Hill, W., & Kraythwohl, D. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*, New York: Longmans Green.
- Chase, W. C., & Simon, H. A. (1973). Perception in chess. In W. G. Chase (Ed.), *Cognitive Psychology*, 4, 55-81.
- Clear, T. (2005, Jun). Comprehending Large Code Bases - The Skills Required for Working in a "Brown Fields" Environment. *SIGCSE Bulletin*, 37, 12-14.
- De Raadt, M., Hamilton, M., Lister, R. et al. (2005). Approaches to Learning in Computer Programming Students and their Effect on Success. *Proceedings of the Annual International Conference of the Higher Education Research and Development Society of Australasia (HERDSA)*. Sydney, Australia. 407-414.

- Fincher, S., (2002). *Experiment kits*, Retrieved August 21, 2006 from the World Wide Web:
<http://www.cs.kent.ac.uk/people/staff/saf/experiment-kits/index.html>
- Fincher, S., Petre, M., Tenenberg, J., Blaha, K., Bouvier, D., Chen, T-Y., Chinn, D., Cooper, S., Eckerdal, A., & Johnson, J. (2004). A multi-national, multi-institutional study of student-generated software. *Proceedings of Kolin Kolistelut - Koli Calling*, Koli, Finland, 20-28.
- Fincher, S., Lister, R., Clear, T., Robins, A., Tenenberg, J., & Petre, M. (2005). Multi-institutional, multi-national studies in CSEd research: Some design considerations and trade-offs. *Proceedings of the First International Computing Education Research Workshop (ICER'05)*, Seattle, USA. 111-121, ACM Press.
- Lister, R. (2004). Objectives and Object Oriented Programming. In S. Mann & T. Clear (Eds.), *Proceedings of the Seventeenth Annual Conference of the National Advisory Committee on Computing Qualifications (NACCO)* (pp. 13-19). Tauranga: NACCO.
- Lister, R., Adams, E. S., Fitzgerald, S., Fone, W., Hamer, J., Lindholm, M., McCartney, R., Mostrom, J. E., Sanders, K., Seppala, O., Simon, B., & Thomas, L. (2004). A Multi-National Study of Reading and Tracing Skills in Novice Programmers. *SIGCSE Bulletin*. 36(4): 119-150.
- Lister, R., Whalley, J., & Clear, T. (2006). *For Discussion: A Framework for a Meta-Project on Students Programmers* (BRACElet Technical Report No. 0106). Auckland: Auckland University of Technology.
- Lister, R., Simon, B., Thompson, E., Whalley, J. L., & Prasad, C. (2006). Not seeing the forest for the trees: novice programmers and the SOLO taxonomy, *Proceedings of the 11th annual SIGCSE conference on Innovation and Technology in Computer Science Education*. Bologna, Italy. 118-122. ACM Press.
- Lister, R., Berglund, A., Clear, T., Bergin, J., Garvin-Doxas, K., Hanks, B., Hitchner, L., Luxton-Reilly, A., Sanders, K., Schulte, C., & Whalley, J. (2006). Research Perspectives on the Objects-Early Debate. *SIGCSE Bulletin*. 38(4).
- McCracken, M., Almstrum, V., Diaz, D., Guzdial, M., Hagen, D., Kolikant, Y., Laxer, C., Thomas, L., Utting, I., & Wilusz, T. (2001). A Multi-Institutional, Multi-National Study of Assessment of Programming Skills of First-year CS Students. *SIGCSE Bulletin*. 33(4).125-140.
- Oliver, D. et. al. (2004): This course has a Bloom rating of 3.9. *Proceedings. of the 6th conference on Australian Computing Education*. Dunedin, N.Z., CRIPT, 57. 227-231.
- Perkins, D., Hancock, C., Hobbs, R., Martin, F., & Simmons, R. (1989). Conditions of Learning in Novice Programmers. *Studying the Novice Programmer*. (pp. 261-279).
- Soloway E., & Spohrer J. C. (Eds.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Petre, M., Fincher, S., & Tenenberg, J. et al. (2003). "My criterion is: Is it a Boolean?": A card-sort elicitation of students' knowledge of programming constructs (Technical Report No. 6-03). Kent, United Kingdom: University of Kent.
- Robertson, J., & Bond, C. (2001). Experiences of the Relation between Teaching and Research: what do academics value? *Higher Education Research & Development*, 20(1), 61-75.
- Robins, A., Rountree, J., & Rountree, N. (2003). Learning and Teaching Programming: A Review and Discussion, *Computer Science Education*, 13(2), 137-172.
- Schnieder , E., & Gladkikh, O. (2006). Designing Questioning Strategies for Information Technology Courses. *Proceedings of the 19th Annual Conference of the National Advisory Committee on Computing Qualifications (NACCO)* (pp. 243-248). Wellington, New Zealand: NACCO.
- Soloway, E. , Ehrlich, K, Bonar, J., & Greenspan, J. (1983.) What do novices know about programming? In B. Shneiderman & A. Badre (Eds.), *Directions in Human-Computer Interactions* (pp. 27–53). Norwood, NJ: Ablex Inc.
- Soloway, E., & Spohrer, J. (Eds.). (1989). *Studying the Novice Programmer*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thompson, E., Whalley, J. L., Lister, R., & Simon, B. (2006). Code Classification as Learning and Assessment Exercise for Novice Programmers. *Proceddings of the 19th Annual*

- Conference of the National Advisory Committee on Computing Qualifications (NACCQ) (pp. 291-298). Wellington, New Zealand: NACCQ*
- Wenger, E. (1998). *Communities of Practice* (1st ed.). Cambridge: Cambridge University Press
- Whalley, J. L., Lister, R., Thompson, E., Clear, T., Robbins, P., Kumar, P. K. A., & Prasad, C. (2006). An Australasian Study of Reading and Comprehension Skills in Novice Programmers, using the Bloom and SOLO Taxonomies. *Proceedings. of the Eighth Australasian Computing Education Conference (ACE2006)* (pp. 243-252). Hobart, Australia :CRIPT.
- Wiedenbeck, S., Ramalingam, V., Sarasamma, S. & Corritore, C.L. (1999). A comparison of the comprehension of object-oriented and procedural programs by novice programmers. *Interacting with Computers*, 11, 255-282
- Zeni, J. (1998). Ethical Issues and Action Research. *Educational Action Research*, 6(1), 9-19.

APPENDIX A

The appendix below provides a useful set of authorship guidelines for MIMN researchers. This has been kindly supplied to the authors by Sally Fincher and is reported as extracted (and slightly adapted) from the British Sociological Association *Authorship Guidelines for Academic Papers* http://www.britisoc.co.uk/new_site/text_index.php?area=publications&id=20, accessed 3 March 2005

Authorship Glossary

Agree early: Agree often: *Authorship should be discussed between researchers at an early stage in any project and renegotiated as necessary. Where possible, there should be agreement on which papers will be written jointly (and who will first author each paper), and which will be single authored, with an agreed acknowledgement given to contributors. Many disputes can be avoided by a clear common understanding of standards for authorship (especially in multi-disciplinary groups). A record should be made of these discussions. Early drafts of papers should include authorship and other credits to help resolve any future disputes.*

Honorary authorship: *named authors who have not met authorship criteria. Deprecated.*

Ghost authorship: *individuals not named as authors but who have contributed substantially to the work. Deprecated.*

Legitimate authorship:

1. *Everyone who is listed as an author should have made a substantial direct academic contribution (i.e. intellectual responsibility and substantive work) to at least two of the four main components of a typical paper:-*
 - a. *Conception or design*
 - b. *Data collection and processing*
 - c. *Analysis and interpretation of data*
 - d. *Writing substantial sections of the paper (e.g. synthesising findings in a literature review or a findings/results section)*
2. *Everyone who is listed as an author should have critically reviewed successive drafts of the paper and should approve the final version.*
3. *Everyone who is listed as author should be able to defend the paper as a whole (although not necessarily all the technical details).*

Non-legitimate authorship claims

The following—by themselves—do not justify authorship

1. *Contribution of ideas*
2. *Paying for the work*

3. *Reviewing the work*
4. *Editing the work*
5. *Acquiring the funding*
6. *Collecting the data*
7. *Supervising the research group*

Order of authors

1. *The person who has made the major contribution to the paper and/or taken the lead in writing is entitled to be the first author*
2. *Those who have made a major contribution to analysis or writing (i.e. more than commenting in detail on successive drafts) are entitled to follow the first author immediately; where there is a clear difference in the size of these contributions, this should be reflected in the order of these authors.*
3. *All others who fulfill the criteria for authorship should complete the list in alphabetical order of their surnames.*
4. *If all the authors feel that they have contributed equally to the paper, this can be indicated in a footnote.*

Acknowledgements

All those who make a substantial contribution to a paper without fulfilling the criteria for authorship should be acknowledged, usually in an acknowledgement section specifying their contributions. These might include interviewers, computing staff, clerical staff, statistical advisers, colleagues who have reviewed the paper, students who have undertaken some sessional work, the supervisor of a research team and someone who has provided assistance in obtaining funding.

Copyright © 2007 Jacqueline L. Whalley, Tony Clear and Raymond Lister.

The author(s) assign to NACCQ and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author(s) also grant a non-exclusive licence to NACCQ to publish this document in full on the World Wide Web (prime sites and mirrors) and in printed form within the Bulletin of Applied Computing and Information Technology.

Any other usage is prohibited without the express permission of the author(s).

Copyright © 2007 NACCQ, Krassie Petrova, Michael Verhaart , Alison Young and Tony Clear. All rights reserved.
Individual articles remain the property of the authors.