THE PEPCAA PROJECT: FORMATIVE SCENARIO-BASED CAA IN PSYCHOLOGY FOR TEACHERS

Victoria Crisp and Christine Ward

The PePCAA Project: Formative Scenario-Based CAA in Psychology for Teachers

Victoria Crisp, University of Cambridge Local Examinations Syndicate, crisp.v@ucles.org.uk;

> Christine Ward, Ward Educational Consulting, ward.education@btopenworld.com

Abstract

The European funded PePCAA project developed a series of scenario-based computer delivered formative assessments of pedagogical psychology for teachers and trainee teachers, using a range of software features, including the addition of confidence measurement. Scenarios were based on realistic classroom situations and focused on problem solving or on utilising best practice. The PePCAA Learning Assessment Circle (PLAC) provided a framework for indexing the kind of processes required of users. Examples of assessments are described, together with results of small scale trialling in the UK. Transfer of assessments between countries proved more difficult than expected. There is scope for further development of the PePCAA approach and for its application in other subjects.

Introduction and project aims

The PePCAA project (Pedagogical Psychology Computer Assisted Assessment), which ran from late 2002 to December 2004, was a European Commission funded project led by FIM – Neues Lernen, a unit within the University of Erlangen-Nuremberg in Germany. Other partner institutions were from Belgium, Spain, Sweden and the UK (see Acknowledgements). The focus of this article is on work carried out in the UK.

The project developed a series of scenario-based computer-delivered formative assessments in pedagogical psychology for teachers and trainee teachers. The project had a two-fold aim, as explained in more detail below: to provide a tool to improve understanding of pedagogical psychology and to explore the potential of more innovative techniques of computer assisted assessment.

The principal project aim was to improve the quality of teaching in the member countries, by linking psychological theory to practical instances of its use in the classroom. The need for this had different origins across the partner countries. In Germany, for example, courses for trainee teachers include a significant component of psychological theory, but do not draw out its implications for classroom practice. In the UK, on the other hand, teacher training is very practical in orientation. Taught courses develop skills such as planning lessons, setting learning objectives and managing a class. Far less time is spent in learning about psychological research and theories relevant to teaching. The PePCAA scenarios are designed to promote interest in pedagogical psychology, to stimulate discussion on classroom applications and to help teachers and trainees to assess their own knowledge of psychology and to develop that knowledge further.

The second main aim of the project was to explore more innovative techniques of computer assisted assessment. Much implementation of CAA to date has used the capacity of the computer to deliver objective testing efficiently, to reduce the burden of marking and to provide fast feedback. Many examples have used multiple-choice and other closed questions; there has been far less use of the interactive potential of the computer. The formative nature of the PePCAA assessments gave the partnership the opportunity to experiment with assessment techniques which would not be sufficiently rigorous in a summative test.

The use of CAA for PePCAA offered motivational benefits. Computer delivered formative testing has previously been shown to have a beneficial effect on learning and on students' results in final examinations, as well as being generally well liked by students (see Bull and Stephens 1999 and, for a more recent example, Pattinson 2004). Many examples, however, have been simply 'practice tests' for the final summative test. For a consideration of the contrast in student reactions between 'improving grades' and 'improving learning', see Baggott and Rayne (2004).

CAA development

The PePCAA project decided at an early stage to buy in software and to concentrate their own efforts on the development of the assessment content. After careful consideration of the requirements of the planned assessments and the features of available CAA software and VLEs (Virtual Learning Environments), QuestionMark's Perception software was chosen. One important consideration was that user prompts were available in all the project languages. FIM purchased and hosted the software and provided IT support.

Software features used in some or all of the PePCAA assessments included:

- a scenario (or 'case study') description at the beginning of each assessment
- the option to read a short explanation of relevant psychological ideas before moving on to the questions (there was no penalty for doing so)

- use of a range of objective question types multiple-choice, multiple-response (where more than one response can be chosen), pull down list (in which a response can be chosen more than once) and matching (similar to a pull down list, but each response may be chosen only once)
- use of 'essay' questions, requiring the candidate to type a response and followed by feedback stating the main points expected
- use of an 'essay' question followed by a multiple-response question asking which of the listed points the candidate had included, with feedback on the choices made
- division of some assessments into blocks, with no possibility to go back to change the answers in the first block; this was used to ensure that answers to essay questions were not changed after feedback
- 'branching' dependent upon the candidate's response to a question (see examples described below)
- addition of confidence measurement (see below) to some assessments; this required some additional programming
- explanatory feedback, often with weblinks, at the end of each assessment.

Metacognition and confidence measurement

Metacognition is knowledge or awareness of one's own knowledge (e.g. Flavell, 1979). Driscoll (1994, p 104) wrote that "metacognition refers to one's awareness of thinking and the self-regulatory behaviour that accompanies this awareness". In other words, metacognition is awareness of one's own knowledge *and* being able to evaluate that knowledge. Metacognition is important, because being able to identify the knowledge in which you are confident allows you to use your knowledge effectively (Hunt, 1993, Leclercq, 1983). Metacognition is also thought to have benefits in learning contexts, since awareness and reflection on current abilities and areas of weakness are likely to improve study decisions. Developing users' metacognitive awareness could contribute to the formative aim of the PePCAA assessment tool.

Various researchers have developed methods of measuring metacognition (e.g. Gardner-Medwin and Gahan, 2003, Leclercq, 1983, Leclercq, 1993, Leclercq and Poumay 2005). Using the technique described by project members Dieudonné Leclercq and Marianne Poumay (2005), the PePCAA project incorporated confidence measurement into some of the scenario-assessments, including one developed in the UK (UCLES Scenario 3). Users are asked to indicate their confidence after answering each automatically scored question.

How confident are you? ○ 0% ○ 20% ○ 40% ○ 60% ○ 80% ○ 100% It is then possible to calculate the following indices for the whole assessment:

Confidence index: average confidence rating on questions answered correctly (a value over 50% is desirable)

Prudence index: average confidence rating on questions answered incorrectly (a value under 50% is desirable)

Nuance: confidence index minus prudence index (a value above 20% indicates the student has a reasonable awareness of his/her own knowledge).

Fred Neumann of FIM developed a workaround, using JavaScript and ASP (Active Server Pages) to allow confidence data to be processed and reported and also integrated an available HTML template engine to give more flexibility in the design of reports.¹

The concept of scenario-based assessments

The assessments developed in this project set a series of questions in the context of a believable classroom scenario. The intention was to base the assessments on practical situations, but with a strong theoretical underpinning. This model views the teachers (or teacher trainees) as decision makers, and judges their psychological knowledge and understanding by assessing the justification they provide for their decisions. A further aim was to develop awareness of why some strategies are more effective than others and to develop user knowledge of relevant psychological theory and research.

Most of the scenarios developed fall into one of two main groups.

- Problem Solving: Scenarios describe a difficulty which has arisen in a classroom and ask users to diagnose causes, to suggest possible solutions and the theoretical basis for them or to predict the effects of a proposed intervention.
- Best Practice: Users are asked to plan the best strategy to achieve a certain goal (e.g. teach number facts in a motivating and memorable way), to justify the strategy by reference to theory and to predict the outcomes.

Each partner institution wrote a number of scenario-based assessments. These were reviewed by at least one other partner institution and then revised. After a number of assessments had been prepared, the PePCAA Learning Assessment Circle (PLAC) was developed to provide a framework for indexing the processes required of users in each scenario and also, a tool for supporting question development in further assessment writing (see Appendix 1). PLAC has four sections representing 'knowledge, understanding and application,' 'demonstrating awareness of required information,' 'using knowledge or understanding to make a judgement' and 'metacognition.' Each section contains several sub-categories of processes required of users. Some questions require more than one of these processes.

¹ It is envisaged that details of this will be added to the PePCAA website before July 2005.

Development of scenario-assessments in the UK

In the UK, scenario-assessments were developed by David Whitebread of the Faculty of Education, University of Cambridge, in conjunction with UCLES. These scenario-assessments were grounded very strongly in practical situations and tried to test relevant psychological theories in an applied way.

Most scenario-assessments begin by describing a situation, for example:

UCLES Scenario 1: History lesson

A class of fourteen year old students appeared bored and uninterested in their last History lesson; in a quick question and answer session at the end of the lesson it was clear they had remembered little and understood less.

In this scenario-assessment, users initially answer questions about the strengths and weaknesses of the strategy adopted by the teacher and are then asked:

Which of the three approaches below would you use to make the next lesson more motivating and effective?

O Put children into teams and award points for best work; you devise a quiz for the end of the lesson with chocolate bars for the best team.

Change the way you present the information to the children. Use a mixture of verbal and visual resources. Find artefacts, videos, pictures and websites and use them to present information to your students.

O Devise a drama about key historical events you want to cover; get children to work in groups, develop and act out the drama; different groups present the same events from the viewpoints of different characters in the drama; at the end of the lesson organise a debate about who was to blame for the events which occurred.

This branches to questions specific to the choice made, asking users to justify their choice and to identify the likely outcomes of other approaches.

A different situation is presented in scenario-assessment 5:

UCLES Scenario 5: dance lesson with 12 year olds This assessment includes seven questions which are separated into two sections. WARNING: During this assessment some marks will not be awarded for correct responses to multiple response questions where too many incorrect responses are also selected. The class are beginning a new unit of work in PE. The unit of work covers dance and movement. The teacher plans for the children to work in mixed sex groups. Each group will design and perform a dance. When introducing the new unit the boys complain that 'dancing is for girls' and they don't want to do 'girls stuff'.

Users are asked to choose a strategy and then to justify it and to identify the likely advantages of the other possible strategies:

3 of 4 The other option that you did not choose was (c) which involved splitting the children into single sex groups and letting them choose what kind of dance to do. Which of the following are likely advantages of this approach? it will encourage the boys to take part because they do not have to do the same thing as the girls it will encourage the boys to take part due to peer influence it will increase the boys' motivation by making the task of interest to them

Some scenario-assessments include direct questions about a relevant theory or questions to address other processes from PLAC.

Users are given feedback for each question, sometimes with scores. For objective questions feedback is specific to the answers chosen, as in this example from scenario-assessment 5:

5 (of 4
	The other option that you did not choose was (c) which involved splitting the children into single sex groups and letting them choose what kind of dance to do.
	Which of the following are likely advantages of this approach? ☑ it will encourage the boys to take part because they do not have to do the same thing as the girls ☐ it will encourage the boys to take part due to peer influence ☑ it will increase the boys' motivation by making the task of interest to them
	2 out of 2 Yes, this method allows the boys to do something different from the girls which may encourage their involvement. Yes, this method will increase the motivation of most of the boys by making the task of interest to them because it now involves superheroes.

Trialling of scenario-assessments in the UK

Each partner institution trialled some scenario-assessments in their own country in order to investigate their value and to inform improvements. Two small trials were carried out in the UK with teacher trainees enrolled on PGCE courses. The majority of participants were females under 25 years of age. Participation in the trialling was voluntary.

Trial 1 – Scenario-assessments about primary pupils

The participants were sixteen students who were training to teach primary aged children. Eight assessment scenarios were trialled, seven written in the UK and one from Sweden. Each participant was asked to complete three scenarios although one participant only completed two scenarios and another completed five in the time available. Each scenario was tested by six or seven of the participants. The scenarios that participants were asked to complete were randomised, as was the order for their completion.

Trial 2 – Scenario-assessments about secondary pupils

The participants were seven students who were training either to teach History to secondary school pupils or to teach middle school pupils (7 to 14 years). Four scenario-assessments were trialled, two written in the UK, one from Germany and one from Belgium. All participants completed all four scenarios. The order in which they attempted the scenario-assessments was randomised.

Feedback questionnaires

In both trials, students completed an online questionnaire after each scenarioassessment. The aim was to gather feedback on the scenarios in order to inform further development. Most questions required students to indicate how strongly they agreed or disagreed with a statement. There were also three open questions inviting comments on what they liked, disliked or would change about the assessments. Three student interviews provided additional feedback.

A note on samples used

It should be noted that the samples cannot be considered representative since participants were from one university. Because participation was voluntary, the samples may have included an unrepresentative proportion of students who were particularly interested in psychology. Also, there were more females than males, which is typical of the general population of primary teachers in the UK, but less so of secondary teachers. Despite this, the trialling findings are likely to provide a useful insight into what teacher trainees would think of scenario-based assessments.

Findings and post-trialling improvements

The responses to the closed questions in the feedback questionnaire are summarised in the following tables along with details of the number of participants attempting each scenario-assessment, the average percentage of marks achieved and the average time taken. The values shown for each statement are the average of:

+2 for strongly agree

+1 for agree

0 for neither agree or disagree

-1 for disagree

-2 for strongly disagree

Hence positive values in the table indicate that most participants agreed with a statement, whilst negative values indicate that most disagreed. The sizes of the positive or negative values represent the average strength of agreement or disagreement.

Scenario:	UCLES 2: PE lesson with 10 year olds	UCLES 3: Number facts	UCLES 4: Assessing Science	UCLES 6: Problem Solving	UCLES 7: Attachment Theory and Parenting	UCLES 8: Friendships	UCLES 9: Play	PGU 7 (Sweden): Teaching conceptual change	Mean
Number of students	6	6	6	7	6	6	6	6	
Average % score for scored questions	64.1	90.5	68.5	63.6	43.2	63.3	88.2	61.1	68.8
Average time taken (minutes)	16:44	12:56	18:31	13:45	29:22	12:57	09:57	13:29	16:19
Knowledge of pedagogical psychology is very important to me	0.5	0.3	0.7	1.0	0.3	0.7	1.0	1.0	0.64
My current knowledge of pedagogical psychology is very good	-0.3	-0.5	-0.5	-0.4	-0.7	-0.3	0.0	-0.3	-0.39
The assessment is easy to understand	0.8	1.2	1.5	0.1	1.0	0.2	1.3	0.3	0.88
The assessment is about a concrete real world situation	1.2	1.2	1.7	0.6	1.5	1.3	1.2	1.2	1.22
The assessment is about a relevant situation for teachers	1.5	1.5	1.5	1.0	1.3	1.5	1.2	1.5	1.36
The assessment is well structured	1.0	1.2	0.8	1.0	0.3	0.7	1.2	0.7	0.88
The assessment is too long	-0.5	-0.5	0.3	-0.3	0.7	-0.8	-1.2	-0.8	-0.33
The assessment is stimulating	0.8	1.3	0.7	0.4	0.8	1.0	1.0	0.7	0.87
Working on this assessment was helpful	1.0	0.8	1.2	0.1	1.5	1.0	1.0	1.3	0.95
The feedback given was helpful	0.7	1.2	1.3	1.1	0.8	1.0	0.7	1.0	0.97

Overview of participants' feedback, Trial 1 - Primary

Scenario:	UCLES 1: history Lesson	UCLES 5: Dance lesson with 11 year olds	Labset 2 (Belgium): Why success isn't explained like failure	FIM 8 (Germany): Self- managed learning	Mean
Number of students	7	7	7	7	
Average % score for scored questions	62.6	67.0	-10.0	32.3	38.0
Average time taken (minutes)	11:01	07:15	08:13	19:06	11:24
Knowledge of pedagogical psychology is very important to me	0.4	0.6	0.4	0.4	0.46
My current knowledge of pedagogical psychology is very good	-0.3	-0.4	-0.7	-0.4	-0.46
The assessment is easy to understand	0.9	0.9	-1.1	-0.4	0.04
The assessment is about a concrete real world situation	1.1	1.4	0.0	-0.4	0.54
The assessment is about a relevant situation for teachers	1.6	1.1	-0.6	0.0	0.54
The assessment is well structured	0.9	1.1	-1.0	-0.4	0.14
The assessment is too long	-0.7	-1.4	-0.4	0.9	-0.43
The assessment is stimulating	0.6	0.7	-0.7	-0.7	-0.04
Working on this assessment was helpful	0.6	0.7	-0.9	-0.3	0.04
The feedback given was helpful	0.0	0.4	-1.0	-0.3	-0.21

Overview of participants' feedback, Trial 2 – Secondary

The responses indicate that most students thought psychology relating to teaching was important, but that they were not confident of their knowledge in this area. The results are very encouraging for the UK scenario-assessments since all values for desirable statements (e.g. '*The assessment is stimulating*') were positive and almost all values for the undesirable statement (i.e. '*The assessment is too long*') were negative. The values suggested that two of the three scenarios from other partner countries were not appropriate for use in the UK, at least not in their current form. This will be discussed below.

For the UK scenario-assessments, the average times taken by users suggest that the length of the assessments was reasonable - long enough to allow users to engage with the material, but not so long that they lose interest. The mean scores were mostly above 60%, suggesting that the assessments were of an appropriate difficulty to challenge users without being inaccessible. The exception among the UK assessments was scenario 7 which was more time-consuming and produced lower scores, both of which could reduce motivation.

The responses to the open questions in the questionnaire also indicated that the UK scenarios were generally very well received and that students seemed to like the concept of such a formative scenario-based assessment tool focussed on psychology relevant to teaching. However, a number of general issues were raised about the tool as well as more specific issues about the individual scenario-assessments. Where possible, changes were made in the light of the findings. A range of views were expressed, by fairly equal numbers of participants, about the types of questions presented.

- Some students liked having open-response questions, even though they were not automatically marked, because they felt this made them really think about the question and formulate their own view. Others felt it was not worth their time composing an answer if it was not going to be marked.
- Some students liked the multiple response questions which sometimes made them think about advantages, disadvantages or outcomes of a teaching strategy that they would not have thought of themselves. Others were critical of the way that marks were assigned to such questions, arguing, with some validity, that there is not one definitive answer to such questions because the topic is somewhat subjective. In addition, some reported that the options in these questions did not always include the answer that they wanted to give.

With such mixed opinions and various advantages and disadvantages of the different question types it was probably beneficial to include a mixture of question types. This should accommodate a range of preferences, encourage the use of a variety of skills and allow both assessment and learning to take place.

Some students said that they would have liked to know at the beginning how many questions the assessment contained. This issue probably arises because the use of question blocks in the Perception software causes the question numbering to restart at the beginning of each block. During posttrialling revisions information about the number of questions and number of sections was added at the beginning of those scenario-assessments where this seemed likely to be helpful.

A few students reported difficulties in understanding some of the psychological language used. Since the aim was to improve their knowledge of relevant psychology, which would involve becoming familiar with such terminology, no post trialling changes were made in the language used. Nevertheless, the students' comments highlight the need for care in the use of technical language in assessments.

In order to make the assessments more formative, and to respond to criticisms that not enough feedback was provided, more links to useful websites explaining relevant psychological topics were added to all UK scenarios after the trialling. These will allow students to follow up areas of interest or topics that they realise they do not currently know enough about.

Transferability of scenario-assessments from one country to another

The initial assumption of the project was that it would be possible to share scenario-assessments between the partners from different countries, subject to translation and with only minor adaptations for language and detail.

Experience from assessment development and trialling showed that transfers between the UK and Sweden worked fairly well, but that transferability between other countries was much lower than expected. There were a number of reasons for this.

- There was a substantial mismatch in the coverage of psychology across the partner countries.
- Even where the initial 'curriculum mapping' suggested that topics and concepts were common across countries, depth of treatment and detail sometimes differed. For example the FIM (Germany) assessment on self-managed learning trialled in the UK went into more depth than would have been expected by UK students.
- The overall approach was much more practical in the UK than in the some other partner countries: the Labset (Belgium) assessment attracted negative feedback when trialled in the UK, as it did not go on to discuss practical advice relating to the theory.
- Assumptions about the structure of education might not fit another country. For example, a scenario from Belgium (not trialled in the UK) assumed that students aged 18 could be choosing whether or not to study science at university, whereas in the UK this choice would determine the subjects studied in the sixth form from age 16. Usually these assumptions were not critical to the scenario and could be changed in editing.
- Assumptions about the pedagogical approach which would be used by the trainee teachers, might not fit another country.
- Translations needed to be carefully edited by a native speaker with access to pedagogical advice, to ensure that the language appeared natural and that the terminology was used correctly.

Conclusion

The PePCAA project has now ended and future development and use of the assessments is unclear. Lack of time and resources prevented some further developments which could benefit the scenario-assessments, for example, the addition of photographs, images or a video clip of a real classroom situation.

It is, however, possible to identify a number of positive outcomes from the work undertaken.

- The project has provided a useful framework for scenario and question development, as well as some finished examples.
- The PLAC (PePCAA Learning Assessment Circle) provides a framework of the mental processes involved in answering questions of this type.
- The good response to the trialling shows the potential of the scenario approach to engage and motivate learners, to improve

their awareness of their own knowledge and areas of weakness and to prompt further learning.

- The approach can also develop students' skills to think about reasons why certain strategies may be beneficial.
- The assessments demonstrate how CAA can move beyond simple multiple-choice, at least in formative assessment.

It is hoped that the PePCAA experience may prompt other research and development, including the application of the scenario-based method to subject areas other than psychology.

For more information about the PePCAA project, access to sample assessments and the full report of the trialling, see <u>http://www.pepcaa.odl.org/</u>

Acknowledgements

The authors acknowledge their debt to all the partners of the PepCAA project:

FIM-NeuesLernen, Friedrich-Alexander Universität, Erlangen-Nürnberg: Walter F. Kugemann (Project leader), Jürgen Großmann, Amrei Tenckhoff, Fred Neumann (IT support), Christine Ward

LabSET - 'Laboratoire de Soutien à l'Enseignement par Télématique', University of Liege, Support Lab for Telematic Learning: Dieudo Leclercq, Marianne Poumay

IN3 - Internet Interdisciplinary Institute, Universitat Oberta de Cataluña (Open University of Cataluña): Carles Fernandez, Elisabeth Galindo

PGU - Institutionen för pedagogik och didaktik, Göteborgs Universitet: Berner Lindström, Marisa Ponti

UCLES - University of Cambridge Local Examinations Syndicate: Alastair Pollitt, Sarah Hughes (to October 2003), Victoria Crisp, David Whitebread (of the Faculty of Education, University of Cambridge)

Institut für Empirische Pädagogik und Pädagogische Psychologie, Ludwig-Maximilians-Universität München: Reinhard Pekrun

The authors also gratefully acknowledge the project funding received from the European Commission under the Minerva strand of the Socrates programme.

References

Baggott, G. and Rayne, R. (2004) *Student perceptions of Computer-based Formative Assessments in a Semi-Distance Module.* Proceedings of the 8th International CAA Conference, Loughborough, pp. 29-31.

Bull, J. and Stephens, D. (1999) The use of Question Mark software for formative and summative assessment in two universities. *Innovations in Education and Training International*, 36, 2, pp. 128-136.

Driscoll, M. (1994) *Psychology of learning for instruction*. Toronto: Allyn & Bacon.

Flavell, J. (1979) Metacognition and cognitive monitoring: A new area of cognitive development inquiry. *American Psychologist*, 34, pp. 906-911.

Gardner-Medwin, A.R. and Gahan, M. (2003) *Formative and Summative Confidence-Based Assessment.* Proceedings of the 7th International CAA Conference, Loughborough, pp. 147-155.

Hunt, D. (1993) Human self assessment: theory and application to learning and testing, in D. Leclercq & J. Bruno (Eds), *Item banking: Interactive Testing and Self Assessment*. NATO ARW, F112, Berlin: Springer Verlag, pp. 177-189.

Leclercq, D.A. (1983) Confidence marking: Its use in testing. *Evaluation in Education,* 6, pp. 161-287.

Leclercq, D. (1993) Validity, reliability and acuity of self-assessment and interactive testing, in D. Leclercq & J. Bruno (Eds), *Item banking: Interactive Testing and Self Assessment*. NATO ARW, F112, Berlin: Springer Verlag, pp. 114-131.

Leclercq, D. and Poumay, M. (2005) Three metacognitive indices of realism in self-assessment, LabSET, University of Liège (available at <u>http://www.labset.net/media/prod/three_meta.pdf</u>)

Pattinson, S (2004) *The Use of CAA for Formative and Summative Assessment – Student Views and Outcomes.* Proceedings of the 8th International CAA Conference, Loughborough, pp. 267-274.

