

Innovation in the Evaluation of Learning Technology

Edited by

Martin Oliver

University of North London

First published 1998 by Learning and Teaching Innovation and Development, University of North London. Printed and bound in Great Britain by the University of North London, 236-250 Holloway Road, London, N7 6PP.

Copyright © 1998 Martin Oliver

All rights reserved. Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, no part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the authors.

The right of the respective authors, named within each chapter of this book, to be identified as authors of this work has been asserted by them in accordance with the Copyright, Design and Patents Act 1988.

ISBN 1-85377-256-9

Contents

Innovation in the Evaluation of Learning Technology	1
Contents	3
Chapter 10 Evaluation of confidence assessment within optional computer coursework	169

Chapter 10

Evaluation of confidence assessment within optional computer coursework

Kim Issroff & Anthony R. Gardner-Medwin

This chapter presents an evaluation of teaching software that employs confidence assessment as a key feature. This is used by medical students at University College London (UCL) for voluntary study and self-assessment in physiology and anatomy. The system (LAPT: London Agreed Protocol for Teaching) requires students to judge their confidence that each of their answers is correct (Gardner-Medwin, 1995). Immediate feedback is given, sometimes with explanations. The evaluation is based on a combination of questionnaire data and usage information gained from use on the UCL campus.

Results from the questionnaire study at the end of the first year medical course (n=136; 65% of the class 95/6) confirmed a high level of voluntary use, particularly towards exam time, and indicated that home use substantially exceeded the recorded use on the UCL campus. Most students (67%) claimed that the confidence assessment was useful to them and that they thought about it for most or all answers. This is borne out by the usage data showing broadly appropriate error rates at the different confidence levels. Forty percent said they sometimes changed their answers as a result of considering their confidence, and this may be an indication that self-explanation is occurring. Many students considered that they were helped in identifying strengths and weaknesses and in distinguishing between knowledge, misconceptions and guesswork.

The combination of questionnaire and usage data provides a clear picture of students' behaviour. There was a generally favourable reaction to confidence assessment as a means to enhance study. The extent of students' preference for home study also has important implications for university strategy and software development.

Introduction

University College London (UCL) has a crowded campus and widely dispersed student accommodation. In this chapter a computer-based

teaching system LAPT (London Agreed Protocol for Teaching) is evaluated. This was set up to encourage effective voluntary computer-based study and self-assessment, particularly for medical students. This initiative originated in the Physiology Department (Gardner-Medwin, 1995), and the period of the evaluation included substantial material used by first year medical students in physiology and anatomy.

LAPT is a package permitting flexible question and answer formats, graphic and dynamic presentations, integration with other forms of CAL, and storage of data and comments. It runs under Windows or MS-DOS, but at the time of this study it ran only under MS-DOS. Further details and downloadable material for evaluation are available from the LAPT web site (see below). With the medical students the largest amount of voluntary study was devoted to mainly text-based MCQ (true/false) material in anatomy and physiology, which is in common with their normal assessment. The key features of LAPT that led to its development at UCL are its incorporation of confidence assessment (see below) and its ease of installation on a home PC-compatible computer. LAPT was available in 1995/6 on over 100 bookable campus computers. Use at home and in student halls was, at the time, entirely on stand-alone PCs using disks created on UCL computer clusters.

The objectives of this study were:

- to investigate some of the general features surrounding the students' use of LAPT
- to investigate the students' attitudes towards the use of LAPT
- to investigate ways in which students' use of LAPT affects their attitudes and work

Particular interest focused on the confidence assessment incorporated within LAPT and on how this affects student learning. The system asks students for answers to questions (which may be a mixture requiring True/False, multiple choice or word and number answers) and then follows this up on each occasion with a request for the level of confidence: 1, 2 or 3. If the student has got the answer correct, this is the number of points awarded (1, 2 or 3). If on the other hand, the answer is wrong then at low confidence (level 1) there is no penalty, while at levels 2 and 3 there is an increasing penalty (-2 and -6 respectively). This non-linear scheme is in a mathematical sense a 'proper' scheme (Gardner-Medwin, 1995): the way to

achieve the highest average score is to have correct insight into the probability of being right and to report this honestly, using level 3 for subjective confidence greater than 80% (odds 4:1), level 2 between 80% and 67% (odds 2:1) and level 1 otherwise. This is clearly explained to the students.

When considering confidence, students can opt to change their initial answer before receiving feedback. Immediate feedback and explanations are given once the confidence has been entered. Final scores are presented at the end of an exercise (a) in terms of this confidence-based scoring system, (b) based on a standard negative marking scheme (+1 or -1) used in their exams at UCL and (c) in terms of percent correct at each confidence level.

Staff observation suggests that students readily understand the basic notion of confidence assessment and relate it to an issue that they perceive as important, that of identifying whether their knowledge is correct or based partly on guesswork. They appreciate why confident wrong answers are so much worse than acknowledged guesses. Confidence data are also useful in the evaluation of course material also. Questions that are identified as eliciting confident wrong answers are particularly important for teachers, since they can pinpoint areas where students have serious misconceptions (Gardner-Medwin & Curtin, 1996). Students are generally well calibrated on average in their subjective confidence judgements (Gardner-Medwin, 1995) though some show systematic overconfidence. The shock of accumulating large negative marks is intended to jolt students into thinking constructively about why they have been mistakenly confident.

The requirement to make a subjective confidence judgment may trigger a range of processes including reasoning, monitoring, reflecting and evaluating. One possible cognitive process of particular current interest is that students may explain to themselves why they think an answer is correct and relate it to a wider range of material. This process is termed 're-explanation' or 'self-explanation' (Chi et al., 1989) and is of considerable pedagogic interest. It can help the student to refine or generalize steps during problem solving. The student may improve his/her understanding by self-explaining aspects of the knowledge domain and/or identifying missing or unreliable knowledge. Self-explanations are thought to be constructive activities that lead to the modification of existing knowledge structures and the construction of new knowledge (Chi & VanLehn, 1991). However, experimental studies usually involve the learner explaining aspects of the domain to either a peer or the researcher. It is not clear whether this is the same as explanation directed at oneself (Ploetzner et al., in press).

Researchers are still working on ways to encourage learners to self-explain, and not all studies on the phenomenon are in full agreement (Barnard & Sandberg, 1996). Confidence assessment may at least sometimes lead students to self explain. This might be expected to benefit their learning.

Methodology

The evaluation involved both questionnaires and usage data. The questionnaires were given to second year medical students in October 1996, regarding their experience in the previous year (October 95 - September 96). The students were given the questionnaires during a laboratory class. The questionnaire included general questions about how much and where they used LAPT, as well as questions about their attitudes towards the various features of LAPT including confidence assessment, final scores and explanations. The students were asked for three reasons why they used LAPT and how their use of LAPT affected their attitudes and their work, in both qualitative and quantitative terms. The questionnaires were coded using an Optical Mark Reader and analysed using Excel.

LAPT runs at UCL on PC clusters under DOS or Windows. Data from student sessions is recorded and collated in three principal ways. Firstly, there is a single line summary for each student's attempt at a particular exercise file, giving the numbers of questions seen and the numbers answered correctly and incorrectly at each confidence level. Secondly, students' volunteered comments on individual questions are recorded in response to encouragement to them to be interactive and say if they think the material is wrong or could be improved. Thirdly, statistics for each individual question identify how many times it has been answered correctly and incorrectly at different confidence levels. For the first two types of data the information was combined and sorted (using Microsoft Excel) according to information entered by the students at the start of a session, giving degree course, year and gender. No specific personal information was recorded during the period of this study unless volunteered by students. Use was optional (with the small exception of introductory sessions) and was encouraged by course organisers as an adjunct to other forms of study. Students were assured that in no way would recorded data be used in their assessment. As revealed in the questionnaire data, much of the use of LAPT took place on students' home computers. No usage data were collected from students' work at home or in halls of residence.

Results

The questionnaire was completed by 136 students (50% male, 49% female, 1 no reply), from the class of 209 students at the start of their second year in October 1996. This was essentially 100% of those who attended a particular practical class. Class time at the start was allocated for completion to reduce any response bias due to variable interest in the questionnaire material. There were no statistically significant differences between male and female responses. Percentages are expressed in relation to the total number of completed questionnaires (136). Blank responses on individual questions were less than 10%.

Students were first asked how much time they thought they had spent using LAPT in the previous year, outside scheduled classes. This varied from none (4%) to >10 hours (36%). Average objectively recorded time on UCL campus machines was 2.8 hours/student. This underestimates the total time spent, however, since much of the student use was off campus (i.e. at home or in residential halls, which were not at the time equipped with networked computer clusters). Sixty three percent of students said that more than half of their use was off campus, while only 18% said they only used the system on campus. From the more detailed breakdown of these data, it appears that about 60% of use was off campus. This surprisingly high figure is significant for future developments since it is clear that students are prepared to go to the trouble of making installation disks for the benefits of working at home.

The students were asked about how easy it was to use LAPT and 96% of the students found the system either easy or very easy to use, with no students saying that they found it difficult to use. Seventy two percent said they generally used LAPT on their own, 25% with one or more friends. When asked which they preferred, the proportions were about the same (67%, 25%).

We asked several questions about confidence assessment. When asked whether they think about the confidence assessment, 63% said they think about it most of the time or every time. One third said they rarely or never think about it. The full breakdown is given in Table 1. The implication is that for a substantial minority of students, confidence assessment is not perceived as relevant or worth the time spent thinking about it. Only 16% however said that they rarely or never paid attention to their final score and confidence breakdown.

Table 1: Students' responses concerning thinking about confidence assessment

Think about confidence assessment	% of students
Every time	17
Most of the time	46
Rarely	25
Never	7
No Reply	5

The students were asked how useful they found both the confidence assessment and the explanations when these were included (Table 2). The explanations were nearly universally valued, with 95% ratings of “useful” to “very useful”. About two thirds of the class gave the same positive ratings for confidence assessment, but 30% rated it less than “useful”. The latter group (40 students) might be expected to correspond roughly to the 32% who said they rarely or never thought about confidence assessment. But in fact 15 of them said they did think about it “most of the time” or “every time”. There was predominant disagreement with the proposition “Thinking about confidence is a waste of time”. On a 5-point scale from *Disagree* (=1) to *Agree* (=5), the average rating overall was 1.9, with 49% on the *Disagree* side and 20% on the *Agree* side of neutral. The proposition was rated higher (3.1) by the 30% subgroup who judged confidence assessment less than “useful”, but still close to neutral (3.0).

Table 2: Responses concerning the usefulness of confidence assessment and explanations

Usefulness	Confidence Assessment % of students	Explanations % of students
Very useful	17	54
	9	12
Useful	41	25
	15	4
Not useful at all	14	1

The students did make substantial use of the different confidence levels. The breakdown of 116,004 responses recorded on campus from this group of students was 65% at C=3 (of which 86% were correct), 21% at C=2 (72% correct) and 14% at C=1 (59% correct). A few sessions (20%) were conducted with completely stereotyped responses or with simulated exam conditions (confidence testing and explanations switched off). In 71% of the sessions where students did vary their confidence assessments, the most frequently reported level accounted for less than 80% of the responses. We conclude that the students were making judgements about whether individual answers were correct, not just broad judgements of mood, and

that on average these judgements were tailored roughly correctly to the probability of being right.

We were interested in whether confidence judgments affected students' answers to the questions, since thinking about how sure one is of an answer can lead a student to change his/her answer. Students rated the proposition "I sometimes change my answer while thinking about confidence assessment". Results are shown in Figure 1. Forty percent agreed to some extent that this was the case, while a further 28% refrained from disagreeing. This suggests that many students do think again about the question at issue and re-explain or attempt to justify their conclusions when faced with the confidence decision, rather than simply acting on how they remember feeling. Future studies could usefully collect data on whether the changes made under these circumstances are random or generally in the correct direction.

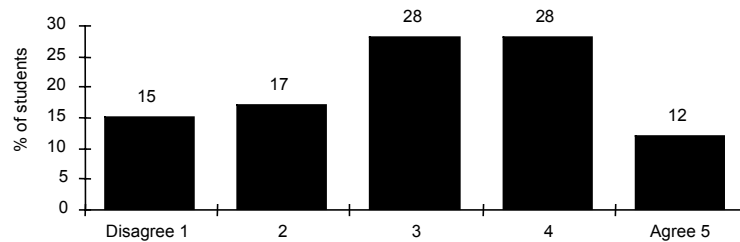


Figure 1: Students' responses to "I sometimes change my answer while thinking about confidence assessment"

Since LAPT was being used voluntarily by the students, we were interested in their motivation for using it. Eighty one percent said they paid attention to their final score and confidence breakdown on an exercise "every time" or "most of the time". Asked whether the results affected their attitude and their work, the breakdowns were 58% yes to 33% no (for attitude) and 49% yes to 41% no (for work). The students were asked how their use of LAPT affected their work and attitudes. Responses about attitude included:

'If high - gives more confidence'

'Helps you decide how much more work and practice needs to be done'

'Useful guide to level of knowledge and real ability'

'Just scares me into working'

'Helps me know how well I understand the subject. If I do well, it helps me to relax and encourages me to do more studying'

'If I do badly, I'll do it again (if there's time) If I do well then it cheers me up!'

Examples of their responses to the question about work are shown below:

'If a low score is given, I know I need to work much harder and go back to the books'

'Try to work harder on areas [in] which score/confidence is low. Revise areas where confidence is weak'

'If I do bad I do more work until I do well'

'If confidence breakdown is good, confidence is increased with the possible drawback of complacency'

Students were asked whether they agreed or disagreed with several propositions about their use of LAPT. These results are summarized in Table 3 giving percentages either side of neutral on a 5-point scale from Agree to Disagree. They show that for about half the students, LAPT helps them to identify when they are guessing, motivates them and improves their confidence.

Table 3: Responses about students' attitudes

	Agree	Disagree
Thinking about confidence helps me identify when I am really guessing.	49%	20%
LAPT helps me understand topics that otherwise I might just learn by heart	29%	31%
Using LAPT motivates me to learn.	54%	15%
I mainly use LAPT for revision.	70%	4%
Using LAPT makes me more interested in the subject.	27%	35%
I feel more confident after I have used LAPT.	60%	12%

It is striking that 70% of students mainly used LAPT for revision. Assessment plays a crucial role in the ways that students work, perhaps particularly medical students. These students were examined in anatomy using true/false questions that have overlapping content with the LAPT material. In the year under consideration, usage data (Figure 2) supported the strong relation to revision and exams. Figure 2 shows clearly that in 94-5 and 95-6 (the year to which the questionnaire data relates) the cumulative

usage rose steeply only at specific times before exams. In the subsequent year successful efforts were made to encourage use as a more integrated part of study, as seen by the more steady rise in the first half of this year.

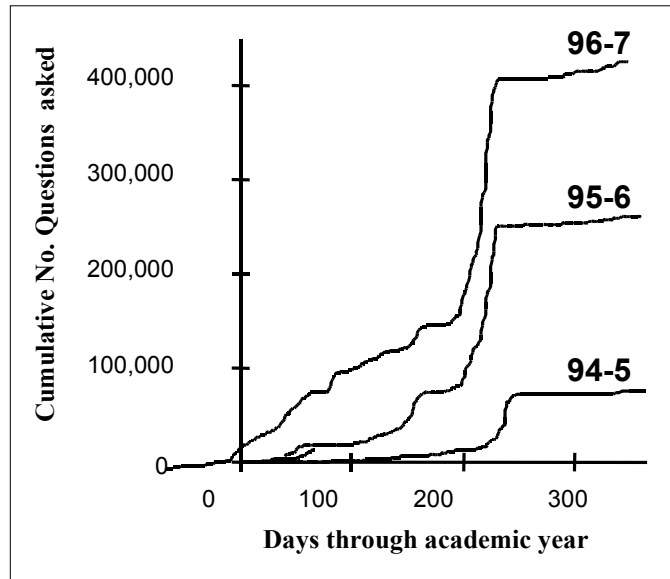


Figure 2: Cumulative use of LAPT recorded on campus at UCL for all students in the 3 years centered on the year of the evaluation study.

An open-ended question about why students use LAPT elicited 259 responses, of which 28% mentioned exams and 20% revision. There were also common references to the explanations, how it was a good way of testing ones knowledge and identifying weak points, and an indicator of how confident one should be. LAPT generally helped to boost the students' confidence. Examples of the responses are shown below:

'Picks out areas where your knowledge is lacking/weak points'

'Very good method of learning, especially some things that may be hard to find in textbooks'

'Gives an indication of how well you actually know things'

'Highlighted difficulties so I knew what topics to revise'

'Because it is easier and more efficient than reading books'

'Because it gives explanations as well as answers'

'Easy and convenient way for testing knowledge as well as gaining knowledge'

Conclusions

Students voluntarily chose to spend a substantial amount of study time with LAPT and more than half of this was off campus. Most of the students reacted positively to the confidence assessment element and there is some evidence that constructive processes occur when they judge their confidence, which may include self-explanations. A significant minority (around 30%) said they did not find confidence assessment particularly useful, though many of these students still claim to think about it while answering questions.

The combination of questionnaire and usage data has been an efficient method of evaluating these students' use of LAPT. The usage data was generated automatically and very little time was spent collecting the questionnaire data from the laboratory class. The Optical Mark Reader minimised the time spent on coding the data, although the open-ended questions were collated by hand, which can be time consuming.

The important finding that over half the students' use of LAPT is at home has implications for student computing policy at UCL. We need to ensure that we can support students' home use of educational software, particularly at a time when resources, both in terms of equipment and space, are limited. This may represent a radical shift for a computing service that has traditionally only provided central support for student computing on campus.

This finding also has implications for development of educational software. This needs ideally to be done in such a way that students can both install and update the software and data from a central source (by disk, CD-ROM, or via the Internet) onto home computers. However, this does complicate the collection of performance and usage data.

The questionnaire data have shown that confidence assessment prompts reconsideration of issues in a question: 40% of students report that they sometimes change their answers as a result. These cognitive processes, perhaps including self-explanation of the reasoning used while thinking

about an answer and the identification of missing or unreliable knowledge, are likely to influence learning. However, we have no data at present about the exact nature of the students' cognitive processes. A relevant factor is that 25% of the students reported working collaboratively. Further studies could investigate the nature of the changes that students make, and ask the the students to retrospectively explain why they made these changes in order to understand the nature of the cognitive processes involved.

The use of questionnaire and usage data in this evaluation has provided a clear picture of many aspects of the students' behaviour and attitudes towards LAPT. The data has both theoretical and practical consequences, favouring the continued support of confidence assessment and computer-based study at home.

Note

The LAPT web site is at <http://www.ucl.ac.uk/~cusplap>

References

- Barnard, Y. F. & Sandberg, J. A. C. (1996) Self-explanations, do we get them from our students? in Brna, P., Paiva, A. and Self, J. (Eds.) *European Conference on Artificial Intelligence in Education*, Ines Mateus, Lisbon, Portugal.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P. & Glaser, R. (1989) Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science* **13**, 145-182.
- Chi, M. T. H., & VanLehn, K. (1991) The content of self-explanations. *The Journal of Learning Sciences*, **1**, 69-105.
- Gardner-Medwin, A. R. (1995) Confidence assessment in the teaching of basic science. *ALT-J (Association for Learning Technology Journal)* **3**, 80-85.
- Gardner-Medwin, A. R. & Curtin N. A. (1996) Confidence assessment in the teaching of physiology. *Journal of Physiology* **494**: 74P
- Ploetzner, R., Dillenbourg, P., Praier, M. & Traum, D. (in press) Learning by explaining to oneself and to others. in Dillenbourg, P. (Ed) *Collaborative learning: computational and cognitive approaches*. Elsevier Press.