AUTOMATED FEEDBACK FOR A COMPUTER-ADAPTIVE TEST: A CASE STUDY

Mariana Lilley, Trevor Barker and Carol Britton

Automated Feedback for a Computer-Adaptive Test: A Case Study

Mariana Lilley, Trevor Barker and Carol Britton
University of Hertfordshire
School of Computer Science, College Lane, Hatfield
Hertfordshire AL10 9AB, United Kingdom
M.Lilley@herts.ac.uk, T.1.Barker@herts.ac.uk, and
C.Britton@herts.ac.uk

Abstract

This paper reports on an empirical study regarding the generation of automated feedback for a computer-adaptive test (CAT) application. In the study reported here, two groups of Computer Science undergraduate students participated in a session of assessment using our CAT application (N=106 and N=82). Participants had 40 minutes to answer 30 questions organised into 5 topics within the Visual Basic.Net subject domain. Participants were provided with feedback on CAT performance via a web-based application specially designed and implemented for this purpose. The feedback provided was divided into three sections: overall proficiency level, performance in each topic and recommended topics for revision. Thirty-one participants from the first group and 25 participants from the second group rated the usefulness of the feedback provided from 1 (not useful) to 5 (very useful). The mean values obtained for the usefulness of the feedback provided were respectively, 4.10 and 3.52. These results were taken to indicate that learners' attitude towards the feedback approach employed was positive overall.

Introduction

Computer-adaptive tests (CATs) are computer-assisted assessment applications in which Item Response Theory (IRT) (Lord, 1980; Wainer, 2000) is employed to adapt the level of difficulty of the test to each test-taker's proficiency level within a subject domain.

At the University of Hertfordshire, a CAT prototype has been designed, implemented and evaluated (Lilley & Barker, 2002; Barker & Lilley, 2003; Lilley et al., 2004a). The CAT application comprises a graphical user interface, an adaptive algorithm based on the Three-Parameter Logistic Model (Lord, 1980; Wainer, 2000) from IRT and a database of questions. All questions in the question database are classified according to topic and Bloom's taxonomy of cognitive skills (Bloom, 1956; Anderson & Krathwohl, 2001).

Findings from previous studies (Barker & Lilley, 2003; Lilley & Barker, 2003) suggested that the CAT approach was a fair assessment method, capable of offering consistent and accurate measurement of learner proficiency levels. Conejo et al. (2000), Fernandez (2003), He & Tymms (2004), Yong & Higgins (2004) amongst others have also reported the benefits of the CAT approach in a wide range of educational contexts.

In spite of the substantial work that has been conducted in this area, research on how CAT proficiency level estimates could be employed to support learners' individual development has received relatively little attention from academic staff and educational researchers.

As part of our research into how such proficiency level estimates can be employed to provide learners with timely and meaningful feedback on their performance, a second version of our feedback prototype has been constructed and evaluated by a group of learners. The first version of our feedback prototype is described in full in Lilley et al. (2004*b*).

Background information on assessment sessions

Two groups of Computer Science undergraduate students, Group 1 (N=106) and Group 2 (N=82), participated in a session of summative assessment using our CAT application.

The assessment session took place in computer laboratories, under supervised conditions. Participants had 40 minutes to answer 30 questions organised into 5 topics within the Visual Basic.Net subject domain. The topics covered in the test are listed in Tables 2 and 3.

Table 1 summarises the performance of both groups of participants. Values for the proficiency level ranged from -3 (lowest) to +3 (highest). In a CAT, the estimated proficiency level for a given learner depends not only on the number of questions answered correctly but also on the level of difficulty of the questions administered during the assessment session. A discussion on how proficiency levels are estimated using IRT is beyond the scope of this paper and the interested reader is referred to Lord (1980) and Wainer (2000). The approach to proficiency level estimate used in the prototype introduced here is described in detail by Lilley & Barker (2002), Barker & Lilley (2003) and Lilley & Barker (2003).

Group	N	Proficiency Level		Correct Responses (%)	
		Mean	Std. Deviation	Mean	Std. Deviation
Group 1	106	0.78	0.78	57.70	9.36
Group 2	82	-0.16	1.16	49.51	9.17

Table 1. Summary of learners' overall performance

Tables 2 and 3 summarise learners' performance according to topic.

GROUP 1	Proficiency Level		Correct Responses (%)	
Topic	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
The Visual Basic.Net Development Environment	0.40	1.68	52.40	24.08
Classes and Controls	0.80	1.45	61.96	19.46
Representing data: Variables and Constants	0.42	1.68	58.86	21.00
Functions and Expressions	0.69	1.67	60.21	21.77
Program Flow	0.73	1.57	55.25	21.76

Table 2. Performance according to topic for Group 1 (N=106)

GROUP 2	Proficiency Level		Correct Responses (%)	
Topic	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
The Visual Basic.Net Development Environment	-0.78	1.63	50.09	21.81
Classes and Controls	0.42	1.32	59.33	20.17
Representing data: Variables and Constants	-0.70	1.88	43.55	21.56
Functions and Expressions	-0.18	1.65	50.33	21.81
Program Flow	-0.19	1.73	44.00	21.32

Table 3. Performance according to topic for Group 2 (N=82)

About the feedback approach employed

It was intended that feedback on assessment performance should be made available to learners via a web-based application. This web-based application is an enhancement of the previous version described in Lilley et al. (2004*b*). The previous version of the feedback made use of an MS Word email attachment to deliver feedback on performance. The current enhanced version was designed to avoid compatibility issues associated with the use of MS Word. It also enabled the number of times learners used the application to be logged.

It was intended that learners should receive feedback on overall proficiency level, performance in each topic and recommended topics for revision.

Overall proficiency level

The overall proficiency level (Lilley et al., 2004*b*) was obtained by employing the Three-Parameter Logistic Model from IRT (Lord, 1980; Wainer, 2000). Figure 1 illustrates how this information was displayed to learners.

Figure 1. Screenshot of overall proficiency level table. Learner's personal details have been omitted

Performance in each topic

The Three-Parameter Logistic Model from IRT (Lord, 1980; Wainer, 2000) was also employed to estimate a proficiency level per topic. An important assumption of our work is that learners' proficiency levels per topic could be mapped into Bloom's taxonomy of cognitive skills (Lilley et al., 2004*b*). For instance, a proficiency level between -3 and -1 would indicate that the cognitive skill *knowledge* has been demonstrated. A proficiency level between -1 and 1 would be taken as evidence that the cognitive skills *knowledge* and *comprehension* have been achieved. Finally, a proficiency level between 1 and 3 would denote that the learner has demonstrated the cognitive skills *knowledge*, *comprehension* and *application*. Higher level cognitive skills are deemed to include all lower level skills. So, a question classified as *application* is assumed to embrace both *comprehension* and *knowledge*.

In the work reported by Lilley et al. (2004*b*), sentences constructed in the light of Bloom's taxonomy of cognitive skills were employed to provide learners with feedback on their performance per topic. Such sentences, however, were deemed 'too long' by some learners who participated in the evaluation of the first version of the feedback prototype. Hence, in the current version of the feedback tool, a pictorial representation was employed instead. This pictorial representation is illustrated in Figure 2.

Performance Summary

This test allows you to demonstrate your competence in the Visual Basic.Net domain at three different levels of learning:

- **Knowledge**, or the ability to recall taught material (e.g. awareness of relevant terminology)
- **Comprehension**, or the ability to interpret and/or translate previously taught material
- **Application**, or the ability to apply taught material to novel situations

Your aim should be to achieve the **Application** level in all topic areas being assessed. The table below summarises **your individual performance** in each topic.

Topic	Knowledge	Comprehension	Application
The Visual Basic.Net Development Environment	✓	✓	✓
Classes and Controls	✓	✓	
Representing data: Variables and Constants	✓	✓	✓
Functions and Expressions	✓	✓	✓
Program Flow	✓	✓	✓

Back to <u>top</u>.

Figure 2. Screenshot of table containing information regarding performance per topic

Recommended topics for revision

Recommended points for revision were extracted from a feedback database, based on the questions answered incorrectly by each learner. Statements in the database comprised directive feedback optionally followed by links to online resources, as illustrated in Figure 3.

Recommended Points for Revision

Did you know this?	Further action that you should do...
When the **ThreeState property** is set to False, the check box can be set to an **Indeterminate** state only through the program and not by any user interaction.	See http://msdn.microsoft.com/library/en-us/cpref/html/frlrfsystemwindowsformscheckboxclassthreestatetopic.asp.
A **Label** control cannot receive focus. Instead it transfers the focus to the control having next higher **TabIndex**. The mnemonic character & (not the character ^) specifies the keyboard shortcut key. The **UseMnemonic** property should be True if the mnemonic character identifies the adjacent character as shortcut key. If **UseMnemonic** is set to False, & will not have any special meaning and will be displayed as it is.	For further information on the **Label Class**, see http://msdn.microsoft.com/library/en-us/cpref/html/frlrfsystemwindowsformslabelclasstopic.asp.
When an exception occurs in a Try block, it will search for a matching **Catch** block associated with that **Try** block. In all cases the **Finally** block is executed.	Review Chapter 7 *"Functions and Procedures"* from Essential VB.Net Fast (J. Cowell).
If you pass a variable "By Value" a copy of that variable is made and passed to the called procedure. If you make any changes to that variable in the called procedure the original variable in the calling procedure is not changed.	Review Chapter 7 *"Functions and Procedures"* from Essential VB.Net Fast (J. Cowell). See also http://msdn.microsoft.com/library/en-us/vbcn7/html/vaconArgPassMechanism.asp and read the section "Argument Passing ByVal".
Watching for endless loops is essential when you are writing Do loops.	For demonstration purposes, create your own example of an endless loop.
In Visual Basic.NET, the **FontStyle** enumeration has a **FlagAttribute** attribute that allows bitwise operations on **FontStyle** values. The result of bit-wise Xor operation is 1 only when one and only one of the conditional expressions is True. Thus, the expression **lblSample.Font.Style Xor FontStyle.Bold** will set the **FontStyle** to **Bold** and replace any existing **FontStyle** of lblSample.	See http://msdn.microsoft.com/library/en-us/cpref/html/frlrfsystemflagsattributeclasstopic.asp. See also Practical Sheet #8.
The **Timer Class** provides a mechanism for executing a method at specified intervals.	Write a simple program that uses a Label and a Timer controls on a form. The program should track the time of day in ten-second increments. In the Timer event, the Label's Text property should be set to the current time. See http://msdn.microsoft.com/library/en-us/cpref/html/frlrfSystemThreadingTimerClassTopic.asp.

Back to <u>top</u>.

Figure 3. Screenshot of table containing recommended topics for revision

Although intrinsically related to the questions answered incorrectly by each learner, the feedback sentences did not replicate the questions themselves. The rationale was that providing learners with a copy of the questions and respective key answers would not foster reflection and/or research (Ellis & Ratcliffe, 2004; Lilley et al, 2004*b*). Moreover, it is often argued that increased exposure of questions would jeopardise their use in future assessment

sessions. The possibility of reusing questions is one of the expected benefits of the creation and maintenance of a database of questions (Freeman & Lewis, 1998).

Learners' perceived usefulness of the feedback approach employed

Thirty-one participants from Group 1 and 25 participants from Group 2 rated the usefulness of the feedback provided from 1 (not useful) to 5 (very useful) using the web page shown in Figure 4. Their responses are summarised in Table 4.

Figure 4. Web page used by participants to rate the usefulness of the feedback

Group	1 Not useful	2	3 Useful	4	5 Very useful	Mean
Group 1	0	1	7	11	12	4.10
Group 2	0	1	16	2	6	3.52
Grand Total	0	2	23	13	18	3.84

Table 4. Learners' perceived usefulness of the feedback approach employed

It can be seen from Table 4 that the feedback approach was well received by the group of learners. Participants also entered some free text comments using the text box provided. Although one participant suggested that s/he "was already aware of some of the information on this feedback page", most learners seemed to value the "Recommended Points for Revision" section. The following examples illustrate their views on this section of the feedback page:

- "I found it useful, gave me an idea of what to revise and work harder on."

- "I now know where I'm going wrong and know how to find out more about the areas of which I scored low marks."
- "I rated this as very useful this is because this does not only allows you check your results but this contain enough updated information on required main topics with useful information, where it can be very useful for revisions. "
- "This feedback page is good because it gives you an insight as to what questions you failed on. It also gives you links to pages that can help you with the questions you did not answer correctly. "
- "This is very useful. It is good to know the exact areas in which I need to work harder. "

Summary and Concluding Remarks

This paper focuses on the second version of the automated feedback tool for a CAT application described by Lilley et al. (2004b).

In the automated feedback prototype described here, learners are provided with information on their overall proficiency levels, proficiency levels per topic and recommended points for revision. Overall proficiency level and proficiency level per topic are calculated using an adaptive algorithm based on the Three-Parameter Logistic Model from IRT. Recommended points for revision are selected from a database based on proficiency level estimates and questions answered incorrectly. Individualised feedback on performance is then made available to learners via a web-based application.

Individual automatic feedback for summative assessment is a useful tool for both tutors and learners. We argue that this has been made easier and more informative by the CAT approach. The encouraging results reported here merit further research and we are currently engaged in collecting views from academic staff on the feedback approach introduced here.

References

Anderson, L. W. & Krathwohl, D. R. (Eds.). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.

Barker, T. & Lilley, M. Are Individual Learners Disadvantaged by the Use of Computer-Adaptive Testing? In *Proceedings of the 8th Learning Styles Conference*, European Learning Styles Information Network (ELSIN), pp. 30-39, 2003.

Bloom, B. S. *Taxonomy of educational objectives. Handbook 1, Cognitive domain: the classification of educational goals*. London: Longman, 1956.

Conejo, R., Millán, E., Pérez-de-la-Cruz, J. L. & Trella, M. An Empirical Approach to On-Line Learning in SIETTE In *Proceedings of the 2000 Intelligent Tutoring Systems Conference, LNCS* 1839, pp. 605-614, 2000.

Ellis, W. & Ratcliffe, M. Improving Results with Positive Directed Feedback in Summative Assessments In *Proceedings of the 8th Computer-Assisted Assessment Conference*, 2004.

Fernandez, G. Cognitive Scaffolding for a Web-Based Adaptive Learning Environment In *Proceedings of 2003 International Conference on Web-based Learning, LNCS* 2783, pp. 12–20, 2003.

Freeman, R. & Lewis, R. *Planning and implementing assessment.* London: Kogan Page, 1998.

He, Q. & Tymms, P. The Development of A Computer Assisted Design, Analysis and Testing System for Analysing Students' Performance In *Proceedings of the 8th Computer-Assisted Assessment Conference*, 2004.

Lilley, M. & Barker, T. Comparison between computer-adaptive testing and other assessment methods: An empirical study In *Research Proceedings of the 10th Association for Learning and Teaching Conference*, pp. 249-258, 2003.

Lilley, M. & Barker, T. The Development and Evaluation of a Computer-Adaptive Testing Application for English Language In *Proceedings of the 6th Computer-Assisted Assessment Conference*, pp. 169-184, 2002.

Lilley, M., Barker, T. & Britton, C. The development and evaluation of a software prototype for computer adaptive testing. *Computers & Education Journal* 43(1-2), pp. 109-122, 2004a.

Lilley, M., Barker, T. & Britton, C. The generation of automated student feedback for a computer-adaptive test In *Proceedings of the 8th Computer-Assisted Assessment Conference*, 2004b.

Lord, F. M. Applications of Item Response Theory to practical testing problems. New Jersey: Lawrence Erlbaum Associates, 1980.

Wainer, H. Computerized Adaptive Testing (A Primer). 2nd Edition. New Jersey: Lawrence Erlbaum Associates, 2000.

Yong, C. F. & Higgins, C. A. Self-assessing with adaptive exercises In *Proceedings of the 8th Computer-Assisted Assessment Conference*, 2004.