Technical Report No: 2001/02

How Students learn to program: Observations of practical work based on tasks completed

Pete Thomas
Carina Paine

2001

Department of Computing
Faculty of Mathematics and Computing
The Open University
Walton Hall,
Milton Keynes
MK7 6AA
United Kingdom

http://computing.open.ac.uk

How students learn to program:

Observations of practical work based on tasks completed

Pete Thomas and Carina Paine

Abstract

Students on a distance education course in introductory object oriented programming are asked to engage in a number of practical, computer-based activities. Each practical activity consists of a significant number of small tasks packaged together into sessions. This paper analyses students' attempts at the tasks in terms of the number of tasks attempted and the time taken to complete the activities with a view to identifying patterns of behaviour. The observed student behaviour is then compared with the behaviour recommended by the course designers. The results reveal both expected and unexpected behaviours and provide some useful feedback on the design of practical activities. In addition, since students perform their practical activities at home on their own PCs, a software recorder gathered the basic information for analysis by another item of software known as the Tasks Completed Tool. The amount of data collected is too large for manual analysis making an automatic analysis tool essential. The paper describes the analysis tool and illustrates how it has been used to identify student behaviours.

Introduction

The AESOP project [4] is an on-going research programme into how students learn to program. A development of the project has been an electronic observatory that allows researchers to observe how students tackle their practical activities as part of their undergraduate studies in computing [9]. Students are taught object-oriented programming in Smalltalk using the LearningWorks software system [8] that has been adapted for Open University use [7]. The adapted system uses a book metaphor in which the software presents itself to the student as sequence of LearningBooks each of which contains a set of exercises (practicals) that ask the student to do a number of computer-based tasks ranging from interacting with a microworld (a simulation) to developing small programs.

The electronic observatory is an extension of the LearningWorks system that records, in a text file, the actions that a student takes when interacting with a LearningBook. A recording consists of a series of time-stamped events that can be replayed at a later date using another piece of software, the Replayer, developed specifically for the project. A third item of software, the Analyser, comprises a number of tools each of which performs an analysis of the data contained in the recordings. Detailed descriptions of the recorder and replayer components can be found in [1, 3].

In this paper we discuss the Analyser's *Tasks Completed Tool* which examines the amount of practical activity a student carries out in a LearningBook. The motivation for this research was to investigate the extent to which students complete the practical work set for them. While the design of the course is such that practical activities form a significant component, anecdotal evidence from other computing courses suggests that students do not complete all practical activities set. This is particularly the case when the available time for study comes under pressure, that is, students abandon practical work in favour of studying the core written materials. Therefore, we wished to find out the extent to which students attempted the practical activities and whether we could elicit any further information about students overall behaviour relating to their practical work.

The *Tasks Completed Tool* is used to identify the times that students spend carrying out practical activities and which activities they perform. To evaluate the effectiveness of the tool and to devise suitable measure to be used in a large-scale investigation, a small-scale trial was conducted in 1999 in which the tool was used to extract data from a set of recordings provided by 30 students working on 8 LearningBooks. A number of different comparisons have been performed on this data which have identified some interesting student behaviours that suggest that all students make comprehensive attempts at the practicals but there are limitations on how much time students are prepared to spend on this work.

Overview of the Tasks Completed Tool

Each LearningBook, which we henceforth abbreviate to LB, is divided into a number of *sessions*, which are subdivided into *practicals* and their associated, but separate, *discussions*. LearningBook practicals ask students to perform 'small grained' programming activities, *tasks*, for example, entering a fragment of Smalltalk code for the system to evaluate. A *discussion* includes solutions and explanations of the tasks set in a practical. Figure 1 illustrates the relationship between, LearningBooks (LBs), sessions, practicals and tasks. The figure shows that there is a relationship between practicals and discussions in that students are encouraged to attempt a practical before turning to the associate discussion.

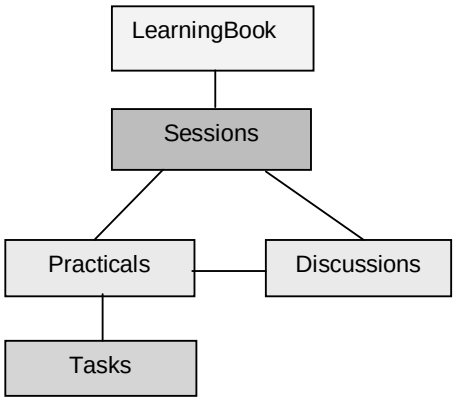

Figure 1 The structure of a LearningBook

The *Tasks Completed Tool* analyses recordings of students' practical activities. The recordings, whilst small in terms of file size, contain a great deal of information which makes them unsuitable for extensive human evaluation. As we were aiming for a large-scale trial that would involve several thousand recordings, an automatic means of analysing them was required. Therefore, the *Tasks Completed Tool* compares the tasks that students have completed successfully (tasks completed) with the tasks they were asked to complete (tasks set) in a particular practical activity. In this context, a *task* is anything a student is asked to do at any instant, and can include evaluating a line of Smalltalk code, creating a new method, amending an existing method, and inspecting variables.

The tool also enables an investigator to identify at which point in a LearningBook students carry out a task; whether they complete a task in a practical, or whether they go straight to the discussion and then carry out the task. The *Tasks Completed Tool* can be used in conjunction with the analyser's time tools [2] to look for relationships between the number of tasks a student has carried out within a LearningBook and the total time taken.

In order to specify the tasks contained in a LearningBook, we adopted the notion of an *ideal recording* – a complete attempt at the activities in a LearningBook – that would be used as a benchmark against which student attempts would be measured. An ideal recording was created by a member of the research team following the instructions in each practical precisely[1]. This information was then used to create a *specification* of the tasks to be completed in each practical.

The tasks completed tool makes a comparison between the specification and each student recording. The output from this tool gives the number of tasks a student was asked to complete in a practical or discussion, and the number

[1] There is an issue as to whether an ideal recording truly represents a practical activity because it depends upon the research worker interpreting the instructions 'correctly'. We attempted to take this variability into account by asking several workers to produce their own ideal recordings and then examined the discrepancies between them. Not only did this provide suitable ideal recordings; it also highlighted possibly ambiguous instructions that the authors of the practical activity could revise.

of tasks they completed successfully. Figure 2 shows an example of the output from the tool when analysing LearningBook 09, session 1, practical 2.

Chapter 09 session 1 practical 2

Number of set tasks = 5

Set tasks:

```
gribbit up, 2 colour, gibbit right, gribbit hone, htonnoy right
```

Number of tasks completed = 4

Tasks completed:

```
gribbit up, 2 colour, gribbit hone, htonnoy right
```

Figure 2 Example output from the Tasks Completed Tool

The tasks set and the tasks completed are listed in the output for ease of comparison. The tasks can be completed by the student in any order, but the tool will only recognise tasks completed in the practical specified.

Some tasks can be identified as completed without dispute in situations where the comparison between what the student did (recorded as a string) and the task specification (another string) is an exact match.

Many tasks cannot be identified as easily. These tasks include evaluations of Smalltalk expressions and the construction of Smalltalk methods, where there may be any number of legitimate differences between a student recording and the specification. As a first step, we use wildcards in the specification allow for some of these differences, including differences in white space, argument values, and the naming of instance variables. For example, if a student were asked to create an instance of the class `Frog`, the specification would be, `*:= Frog new,` making allowance for the student to use any variable to refer to their `Frog` instance.

In some practical activities, students are asked to investigate or experiment where they are expected to devise their own solutions to problems. Again, the use of wildcards in the specification can identify some of this extra work. For example, if students are asked to experiment with the `colour:` method the specification would be `*colour:*.` In effect, the specification simply demands that the string `colour:` should appear somewhere within the student attempt.

Clearly, this scheme does not allow for all possibilities, but the nature of the practical activities, particularly at the start of the course, is often one in which the expected responses are heavily constrained and the number of possibilities is quite limited.

Research questions

Our first question was to ask whether or not students completed all the tasks set and, if not, was there any pattern in their behaviour. We had noticed, from an initial scan of some recordings that students did not seem to complete all the tasks set and, in LBs with multiple sessions, tended to try the first session and then either give up or attempt all sessions in the LB.

Table 1 illustrates the kind of data that the Tool provides. It shows the percentage of tasks completed for a sample of students in LB 09 (the first LearningBook containing any meaningful programming tasks). LB 09 is split into 4 sessions with a total of 21 practicals and 83 tasks. Clearly, the percentage of tasks completed in each session decreases as the student progresses through the LearningBook.

The behaviour of students on other LBs is not so dramatic, but some patterns have emerged.

% Tasks Completed (LearningBook 09)				
Session 1 Practicals 1 - 6 30 Tasks	**Session 2** Practical 7 - 9 15 Tasks	**Session 3** Practicals 10 - 15 21 Tasks	**Session 4** Practicals 16 - 21 17 tasks	**Complete LB** Practicals 1 - 21 83 Tasks
75.00%	65.56%	62.70%	54.90%	69.28%

Table1: The percentage of tasks completed in LearningBook 09

Table 2 shows the average percentage of tasks completed for all LearningBooks analysed (we looked at only those LearningBooks that asked students to perform some programming activities). Typically each LearningBook is expected to be studied over four hours, but for a few LearningBooks less time is allocated.

Learning Book	**Time allocated (hours)**	**No. of Sessions**	**No. of Tasks**	**No. of Students**	**Average % of Tasks Completed**	**Standard Deviation**	**Min % of Tasks Completed**	**Max % of Tasks Completed**
9	4	4	83	6	69.28	26.136	25.3	92.77
10	4	5	43	7	90.37	3.288	79.07	100
12	1	1	57	8	78.95	18.693	0	100
13	4	3	66	13	90.33	6.627	63.64	100
14	4	3	72	14	89.48	5.515	70.83	98.61
15	2	3	58	15	92.64	3.712	82.76	100
16	3	3	33	16	82.01	3.890	45.45	93.94
20	4	1	49	8	64.29	11.514	24.49	91.84

Table 2: Summary data of tasks completed in each LearningBook

Some of the information in Table 2 is shown graphically in Figure 3.

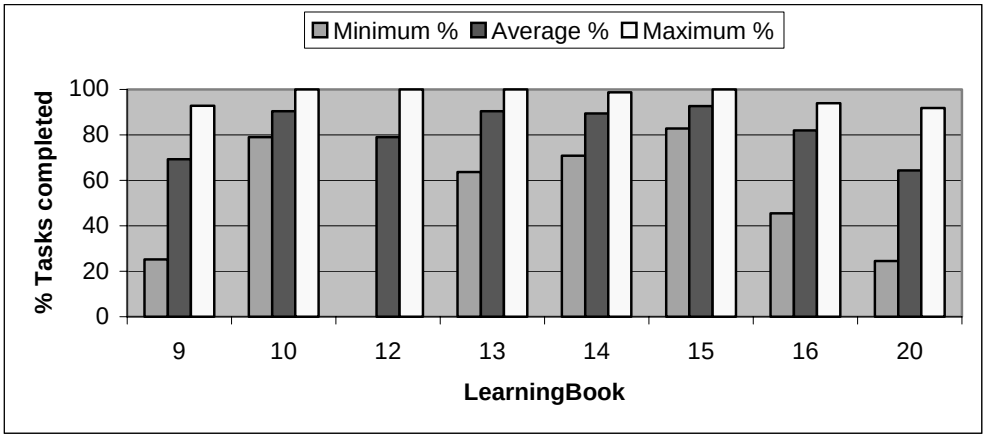

Figure 3 Summary data of tasks completed in each LearningBook

Two features of Table 2 and Figure 3 are immediately apparent. First, on average, students complete over 60% of the tasks set in all LearningBooks. Second, while some students (under 10%) complete all the tasks set, in general the majority of students do not do everything in a LearningBook (80% completed, on average), and we are interested in trying to find out why.

Two natural questions to ask are:

Does the absolute number of tasks set in a LearningBook affect the amount of tasks a student completes in a LearningBook? That is, is the setting of more tasks off-putting to students? Perhaps the apparent volume of work puts students off.

> *We hypothesise that the more tasks there are to do in a LearningBook, the number of tasks completed in the LearningBook will be lower.*

Is the time spent studying a LearningBook related to the number of tasks set in the LearningBook? It would seem obvious that the length of time spent studying would be related to the amount of work set, but we wondered whether the perceived amount of study required affected the students' willingness to complete all the study set.

> *We hypothesise that the time taken to study a LearningBook is proportional to the number of tasks in the LearningBook.*

Throughout the investigations we have examined both the number of tasks set as well as the number of tasks completed (as reported by the *Tasks Completed Tool*) in order to provide a limiting case as follows. The current implementation of the tool detects tasks successfully completed but not tasks attempted. That is, we define an *attempt* at a task as any effort to do the task whether successful or unsuccessful. If an attempt is successful we say that the task has been completed. The tool is designed to identify all successfully completed tasks that match either the expected outcome or an outcome that is close to the expected one (as will be described later). Thus, the tool underestimates the number of tasks attempted. We can therefore say that the number of tasks completed as reported by the tool provides a lower bound on tasks attempted. However, since it is not possible to attempt more than all the tasks set, looking at tasks set provides an upper bound on tasks attempted. We discuss the design of the tool in more detail later in the paper.

Results

LearningBooks

Starting with the first question, we looked at the number of tasks set, and the number of tasks completed in each of 8 LearningBooks. Figure 4 shows a diagram of the number of tasks completed plotted against the number of tasks set. The upper line indicates the maximum number of tasks that could be completed (the number of tasks set). A regression line for the data is also given.

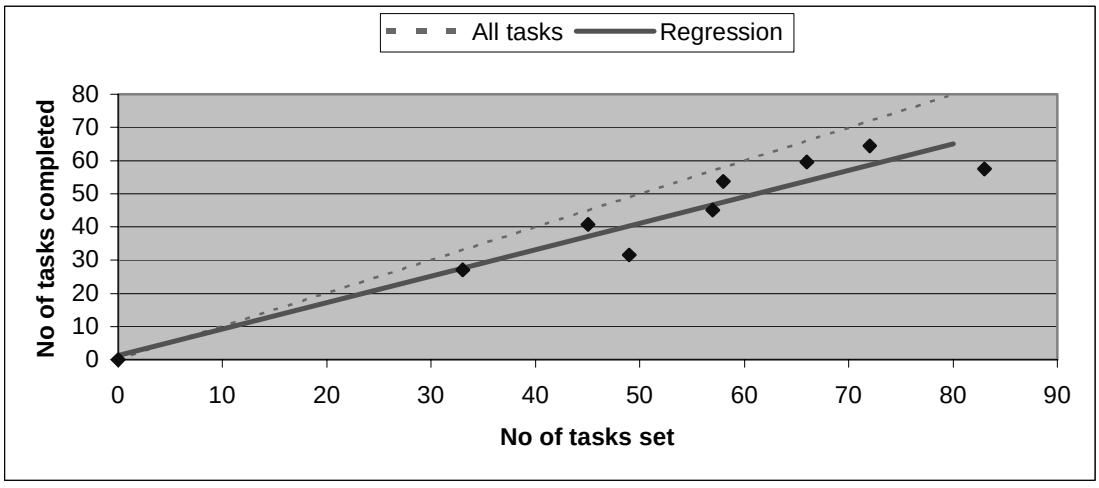

Figure 4 The number of tasks completed compared with the number of tasks set in eight LBs

Spearman's *Rho correlation test* applied to the data of Figure 4 yields a value for ρ_s of 0.9047, when the critical value for ρ_s for a two-tailed test when N=8 is 0.738 (95% confidence) confirms that there is a correlation between the number of tasks completed and the number of tasks set. The regression line has a slope of 0.796 which suggests that, across all LBs, students complete around 80% of tasks set, as observed earlier.

We then went on to compare the number of tasks set with the time spent in a LearningBook, to answer Question 2. Figure 5 shows a plot of the data in which the number of tasks set and the time spent (seconds) is shown for each LearningBook.

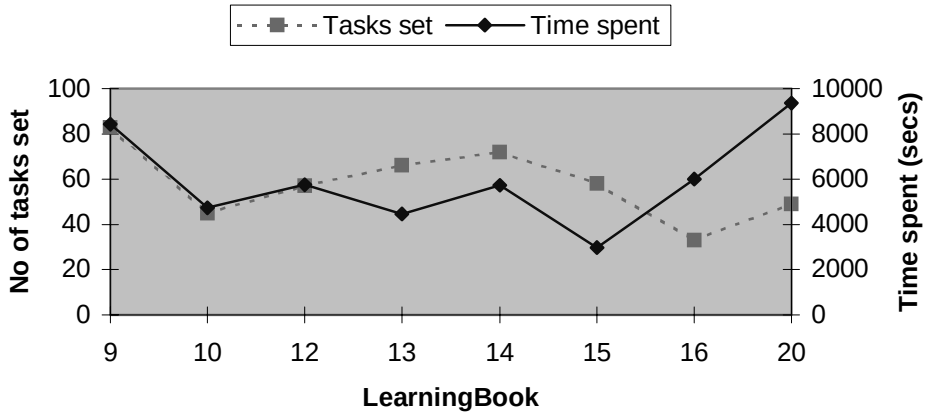

Figure 5 The amount of time spent and the tasks set in each LearningBook

No significant correlation was found between the number of tasks set in a LearningBook and the time spent in a LearningBook. However, looking at the tasks completed as a percentage of all tasks set against the time spent in each LeaningBook reveals a different picture, shown in Figure 6.

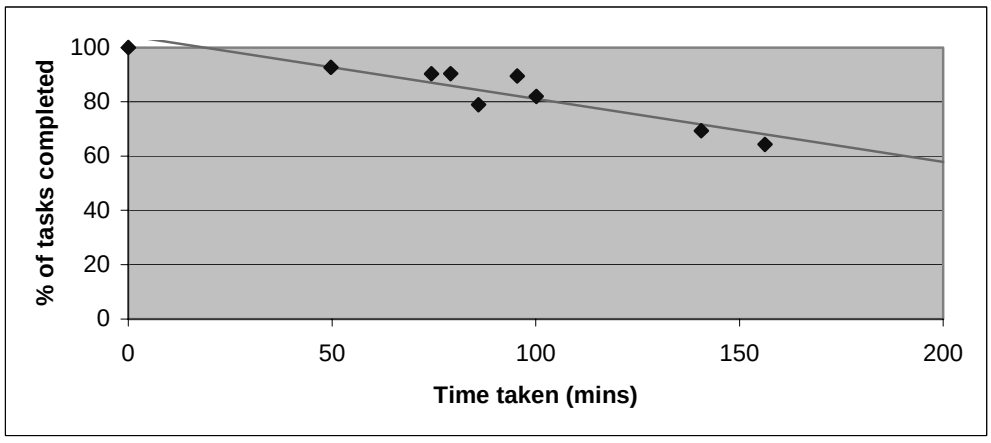

Figure 6 The relationship between the percentage of tasks completed and time spent in each LearningBook

This time Spearman's Rho correlation test shows a strong negative correlation (ρ_s = -0. 9048), when the critical value for ρ_s for a two tailed test when N=8 is 0.738 (95% confidence)). This suggests that the more time a student spends, the smaller the proportion of tasks set are completed.

Sessions

A possibly more interesting factor is that a LearningBook is normally split into several sessions and earlier work (see [2]) showed that students tend to complete a session in one sitting, that is, they normally work through a session continuously. Therefore, looking at sessions in a LearningBook might provide a finer grained view of student behaviour.

Returning to Question 1 we looked at the percentage of tasks completed in each *session* of a LearningBook to see whether there was a relationship between the number of tasks set and the number of tasks completed. As was true at the LearningBook level, there was a significant relationship as Figure 7 shows.

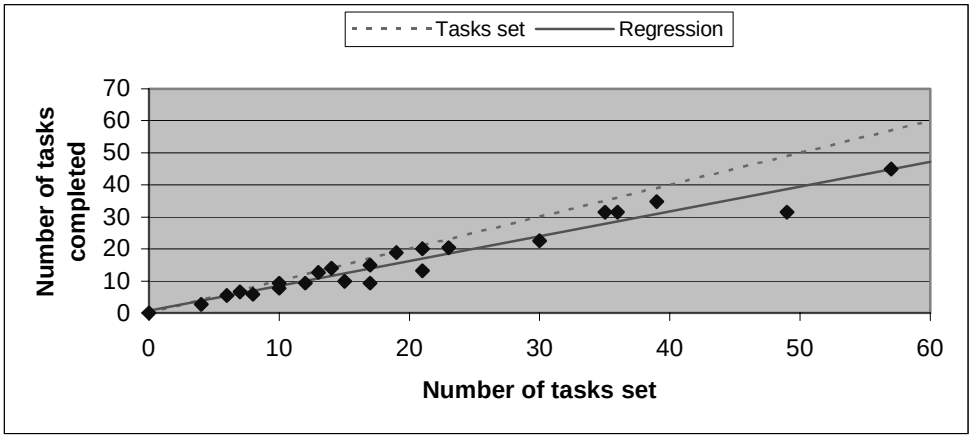

Figure 7 The number of tasks completed in a session compared with the number of tasks set in 8 LBs

The Spearman Rho correlation test gives $\rho_s = 0.9726$, when the critical value for ρ_s for a two tailed test when N=22 is 0.425 (95% confidence), confirming a strong relationship. That is, the number of tasks completed is proportional to the number set with the slope of the regression line being 0.773694.

Turning once again to our second main question, a plot of the number of tasks set against the time spent studying a session is shown in Figure 8.

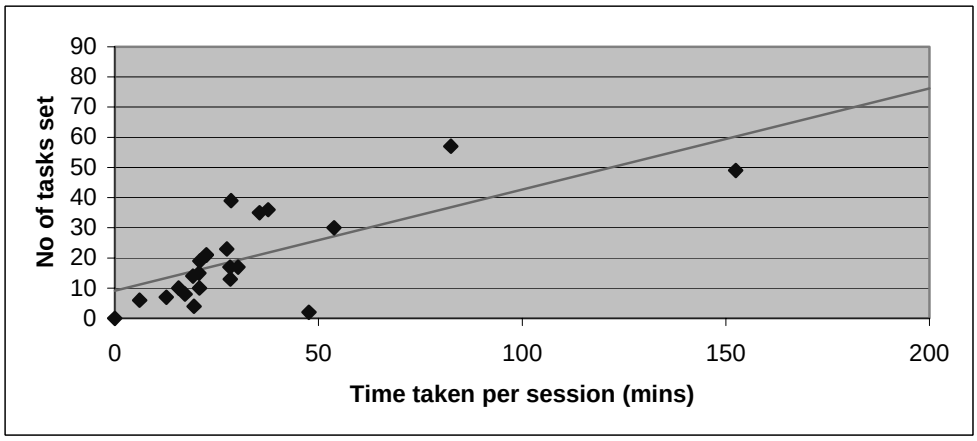

Figure 8 The amount of time spent and the number of tasks set in each session

Spearman's Rho test produces a value of $\rho_s = 0.6897$, when the critical value for ρ_s for a two tailed test when N=22 is 0.425 (95% confidence) confirming a significant relationship. The regression line has a slope of 0.3352. We conclude that, on average for a session, the time taken to study a session is proportional to the number of tasks set which is a relationship that did not hold for LearningBooks.

A closer look at Figure 8 reveals that the two LBs with the greatest value for time taken are LB 12 and LB 20, which are unusual in the sense that they are the only LBs in this investigation that have a single session; all other LBs have two or more sessions. The data of Figure 8 without LBs 12 and 20 is shown in Figure 9 where the regression line has a slope of 0.4951.

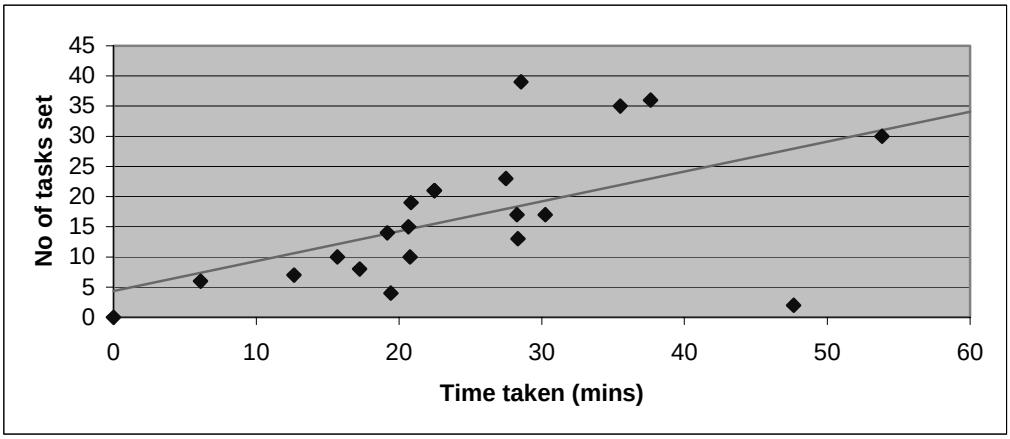

Figure 9 The amount of time spent compared with the number of tasks set in each session excluding LB12 and LB 20

A plot of the number of tasks completed in each session against the time spent by students doing the tasks is given in Figure 10 and Figure 11 shows the data excluding LBs 12 and 20.

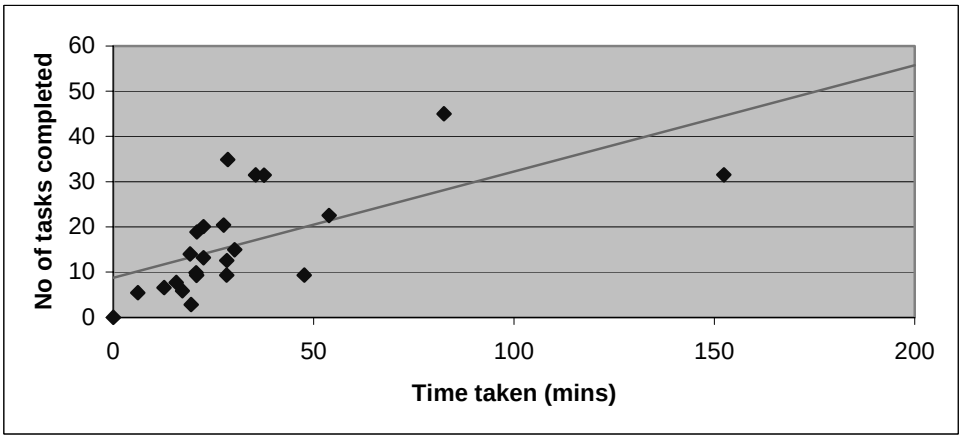

Figure 10 The amount of time spent against number of tasks completed in each session of a LearningBook

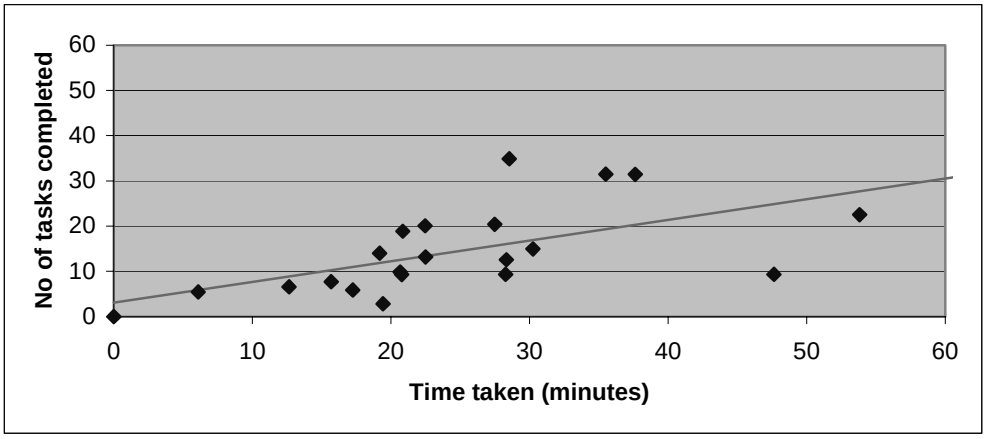

Figure 11 The amount of time spent compared with number of tasks completed in each session excluding LB12 and LB 20

For the complete set of data shown in Figure 10, Spearman's Rho test produces a value of $\rho_s = 0.7708$, when the critical value for ρ_s for a two tailed test when N=22 is 0.425 (95% confidence) confirming a significant relationship. The regression line has a slope of 0.2349. The slope of the regression line excluding LB12 and LB 20 is 0.4571. This illustrates that there is a stronger relationship between LBs with multiple sessions, which suggests that student behaviour when there is only one session in a LearningBook is different from their behaviour when there are multiple sessions.

Task difficulty

The initial analyses discussed above assume that all tasks are of the same complexity either in terms of the time taken to complete or degree of difficulty. However tasks can vary in levels of difficulty, and this is a potentially important point when comparing different LearningBooks. Further analysis attempted to take into account the varying levels of task difficulty. Therefore, we attempted to obtain a view of difficulty by analysing the tasks completed in each LearningBook.

The first step was to define a metric for task difficulty. Initially, we devised three metrics for *task difficulty* as follows:

A the average time (minutes) spent in a LearningBook per task set;

B the average time (minutes) spent in a LearningBook per task completed;

C the recommended time (minutes) to be spent in a LearningBook per task set.

Metrics (a) and (b) are based on student behaviour whereas metric (c) is based upon the course team's perception of the tasks. Values for the three metrics for each of the 8 LearningBooks investigated are shown in Table 3.

LearningBook	Metric A	Metric B	Metric C
9	1.75	2.53	2.89
10	1.84	1.93	5.58
12	1.51	1.91	1.51
13	1.13	1.25	3.64
14	1.33	1.48	3.33
15	0.86	0.93	2.07
16	3.03	3.70	5.45
20	3.19	4.96	3.85

Table 3 Values of the three metrics for task difficulty

Given that the tool that analysed the tasks completed will have, in general, underestimated the number of tasks actually completed, metric A provides a minimum value and metric B a maximum value. Together, metrics A and B provide a range within which the ratio of tasks attempted in the time spent lies.

The data in Table 3 is also shown in Figure 12 which graphically illustrates that the course team generally over-estimates task difficulty (the middle column in each category). It is interesting to note that the two LearningBooks in which the course team's estimate coincides with the actual time taken by students are LB 12 and LB 20.

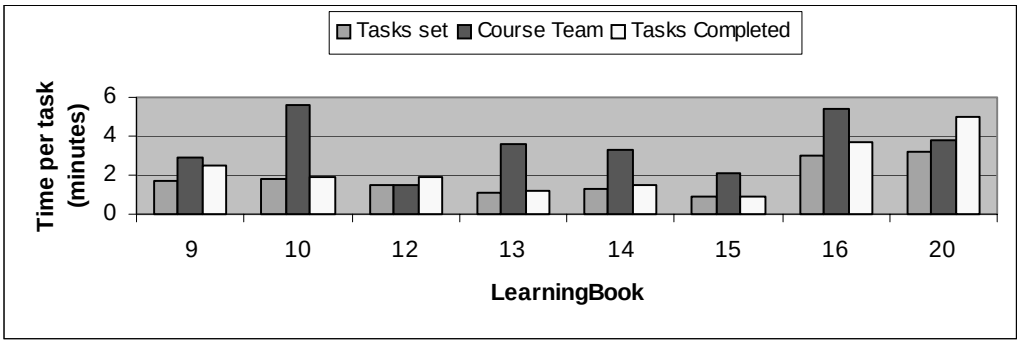

Figure 12 Three measures of task difficulty compared.

The time spent on an individual task averaged over individual learning books shows that, approximately, students spend between 1 and 4 minutes per task whereas the course team felt that they should be spending between 2 and 5 minutes. The difficulty of each session is shown in Figure 13.

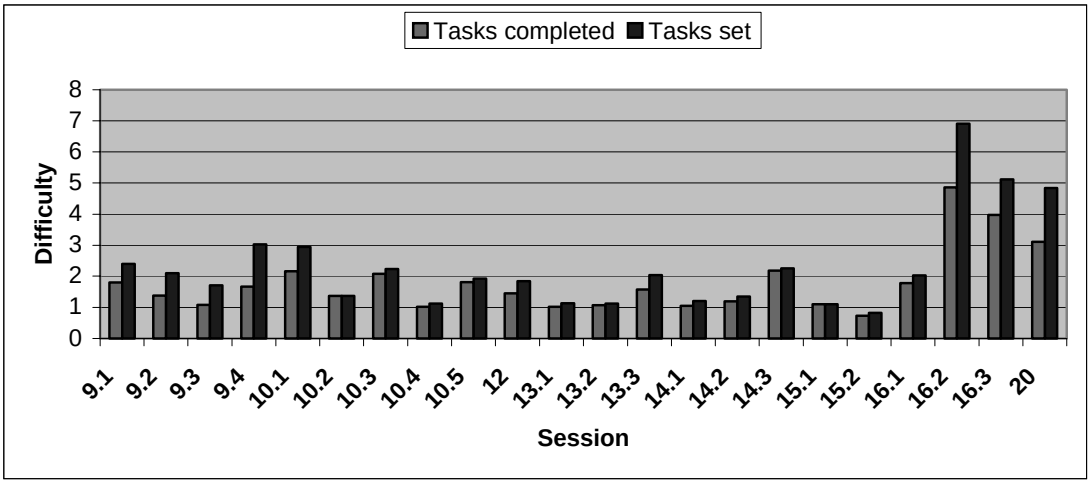

Figure 13 Session difficulty

More revealing is a comparison between the task difficulty and the percentage of tasks not completed in a LearningBook shown in Figure 14.

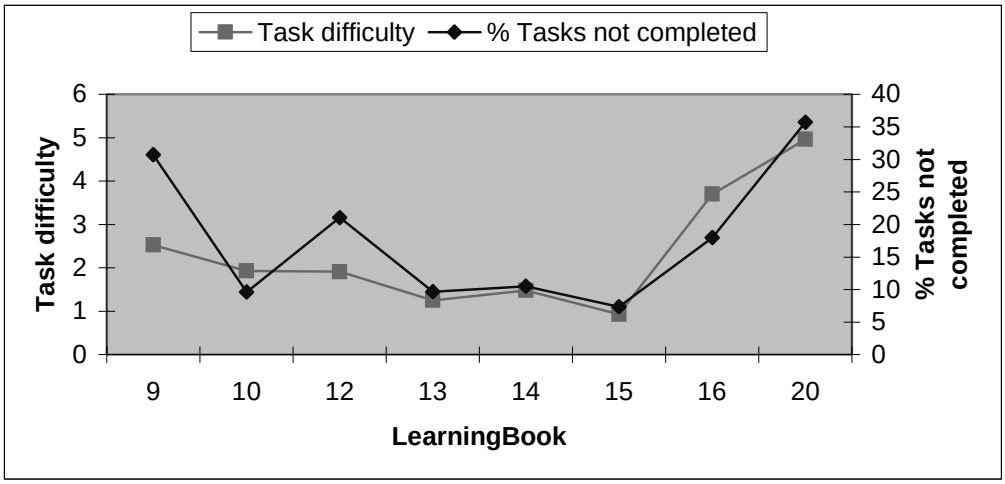

Figure 14 Comparing task difficulty with % of tasks completed in each LearningBook

Spearman's *Rho correlation test* yields a value for ρ_s of 0.762, when the critical value for ρ_s for a two-tailed test when N=8 is 0.738 (95% confidence) confirms that there is a correlation between the number of tasks not completed and the task difficulty.

Sittings

Each session of a LearningBook was designed to be studied in one go. That is, it was expected that a student would complete a session in a single continuous interaction with the computer. A question naturally arises as to the extent to which students followed this pattern given the variation in the number of tasks per session.

If we define a *sitting* as a period of time that a student interacts continuously with the computer without a significant break in activity we can see whether the number of sittings varies with the number of sessions in each LearningBook. The identification of sittings within each recording is not straightforward because a recording is

simply a list of time-stamped events. We cannot be certain whether the time elapsed between two events represents a break in study or an activity, such as reading, which is related to the practical activity being undertaken. Therefore, we analysed the recordings in such a way that any gap between two successive events greater than a certain threshold was taken as a break in study and divides one sitting from the next. For example, a gap greater than one hour would be indicative of a break in study. Figure 15 compares the results of this analysis for three different gap sizes: 30 minutes, 15 minutes and 10 minutes. The figure shows the number sittings identified with the three gap sizes together with the number of sessions in each LearningBook.

For a gap size of 10 minutes or more, the number of sittings exceeds the number of sessions. A gap size of 15 minutes or more usually gives an average number of sittings greater than the number of sessions. A gap greater than 30 minutes results in an average number of sittings that is normally smaller than the number of sessions. The two notable exceptions to this pattern are LB 12 and LB 20 where it is very clear that students do not complete the LB in a single session as recommended, but take 3 or 4 sittings.

In earlier work [2], we took any gap of 10 minutes or more to be an indication that a student has broken off from their studies. This was deduced on the basis that, since the material required to perform individual tasks is contained wholly within the LearningBooks, there is little incentive to interact with other sources of material, and the amount to be studied to perform a task is quite small and only likely to take a few minutes. Our conclusion is that, on average, for LBs with multiple sessions, students either follow the instructions precisely and study each session in a single sitting, or take one or two more sittings.

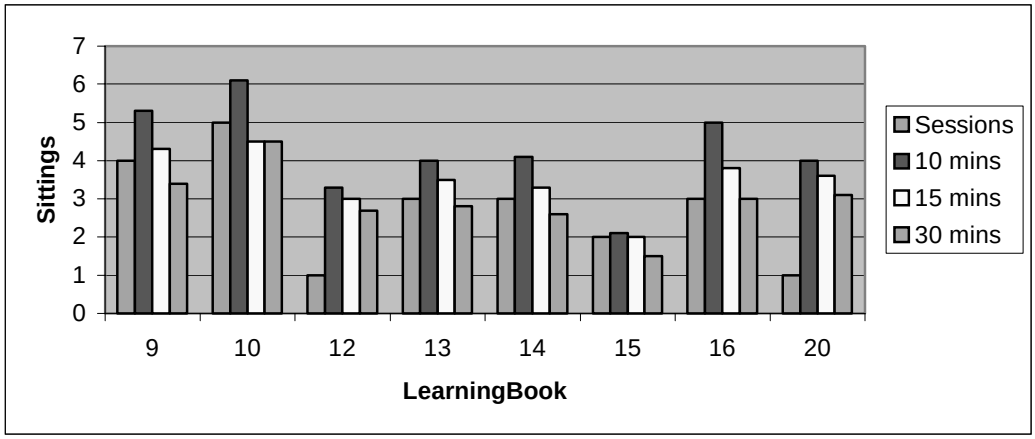

Figure 15 The number of sittings per LearningBook

Figure 16 plots the time per sitting with the same three indications of study breaks. The indications here are that there is a limit to the amount of time students are prepared to spend in a single sitting. For the 10 minute gap, the limit is around 25 minutes, for the 15 minute gap, the limit is around 30 minutes. LearningBook 20 is the exception, but even here the limit is around 40 minutes.

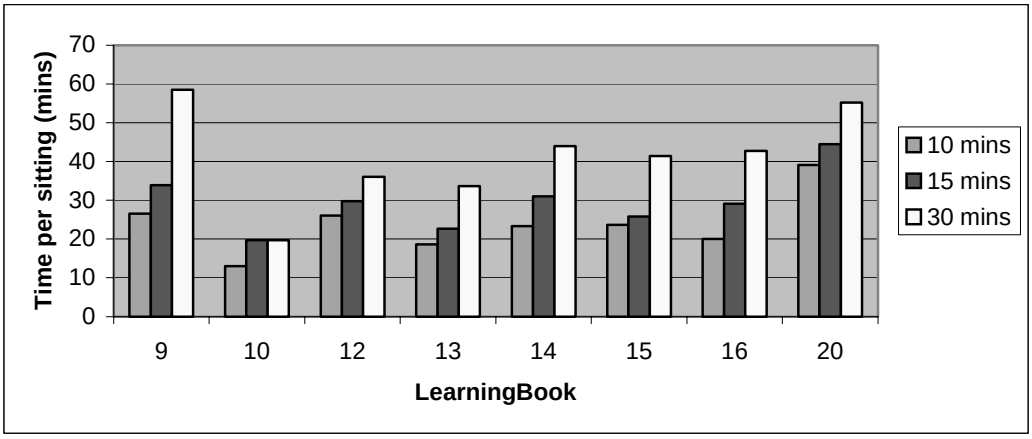

Figure 16 The time per sitting in each LearningBook

It seems reasonable, therefore, to suggest that sessions should be studied in a single sitting designed to take not more than 30 minutes.

The limitations of the tasks completed tool

The version of the tasks completed tool reported upon in this paper has a number of weaknesses. In its initial implementation, the tool looked for an exact match between the specification and the student recording, but has been extended to allow for some variations in student responses. Nevertheless, the tool does not identify all attempts at tackling a task. Where possible we have made allowance for this deficiency by looking at tasks set. Nevertheless we wish to improve the tools accuracy and this is described in more detail in a later section.

A further limitation of the initial implementation of the tool is that it analyses student activity whilst they were working with the practical activities. There is evidence to suggest that some students perform the tasks having looked at the discussions where solutions to the tasks are given. Thus, in some cases, the evidence of tasks being completed are located in a different part of the recordings, though we believe this to be a minor effect.

In some practicals, students are asked to complete the same task more than once, but the tool does not deal satisfactorily with any line of code appearing more than once in a practical.

Clearly, the major weakness of the present tool is that it does not identify whether any unsuccessful *attempts* were made at a task. If a student has not completed a particular task we can look at the practical in more detail to determine whether they have made any kind of attempt at the task.

However, the task difficulty metrics indicate that the tool provides a good estimate of the tasks attempted which gives us reasonable confidence in the conclusions we are drawing. Nevertheless, we wish to improve the tool and the path we are taking is described in the next section.

Future work

Experiments

The data presented here was collected during the 1999 presentation of the course and has enabled us to clarify how the student recordings can be analysed by identifying appropriate questions and developing analytical tools. We collected data from a relatively small number of students: 30 students took part in the experiment but each one did not provide us with a complete set of recordings for the 8 LearningBooks we investigated. Therefore, we have repeated the experiment in 2000 with around 200 students actively taking part and are currently processing the data.

An area that we wish to investigate in more detail is how students tackle their assignments. The LearningBooks we have investigated so far have asked the students to engage with a series of well structured but small-scale tasks. In their assignments, students have more freedom of choice and it will be interesting to see how the analyses we have performed so far compare with student behaviour in a less constrained environment.

Analyses from this year's study will provide more detail about tasks *attempted* by students. Investigations of any relationships between the amount of practical work a student carries out in a LearningBook and their resulting TMA and exam scores will also be conducted.

Task description

A second version of the *Tasks Completed Tool*, currently under development, will look for a specified task in a student recording, and employs a scoring system for tasks successfully completed, or tasks partially completed. This new tool will identify any attempts a student makes at a task, wherever carried out in a LearningBook, and will, it is anticipated, recognise if a student devises an alternative solution to the one used in the specification. The aim is to allow an investigator more flexibility in describing a task. Not only is the tool aimed at identifying attempts at carrying out a task, it also tries to define the accuracy of an attempt through a scoring system. This will enable us to examine questions related to how well a student solves a problem and the number of attempts at solving each problem.

Figure 15 shows an example of the interface for the new tasks completed tool. The user enters the number of the chapter/session/ practical /task they wish to investigate in the appropriate fields. For example, in Figure 15 the text pane shows the task description for the first task in chapter 9, session 1, practical 2. The first line in the edit window is an optional comment describing the task the student was asked to complete. In this simple example, the students were asked to evaluate a single line of code, `gribbit up`. The remaining lines display the scores awarded for the evaluation or variations on the evaluation. In this example, a student evaluating `gribbit up` will be awarded 1, whereas a student evaluating `gribbit` (alone or followed by something else other than `up`) will be awarded half.

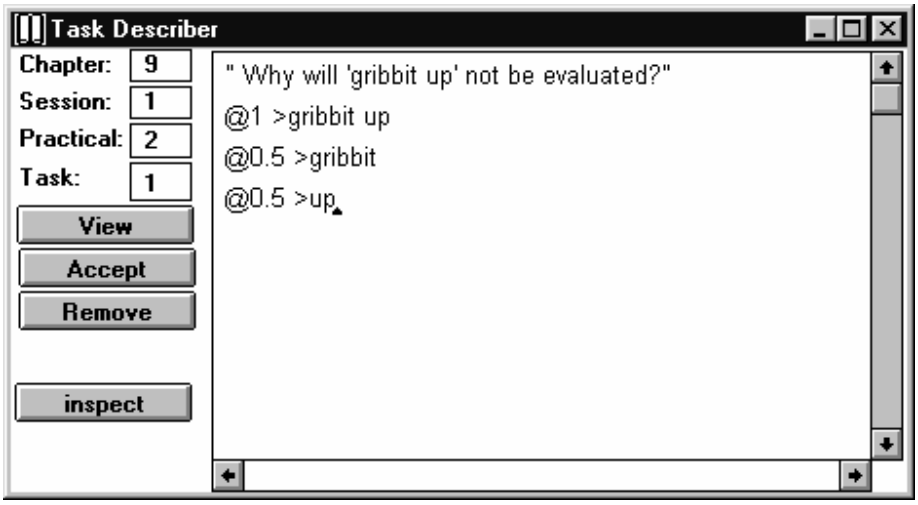

Figure 15 Interface to the Tasks Describer tool

Conclusion

Our initial investigations showed that while some students (under 10%) complete all the tasks set in their practical work, in general the majority of students do not do everything in a LearningBook. Our Tasks Completed Tool showed that on average across all LearningBooks, students complete 80% or more of tasks set.

This encouraged us to pose two main hypotheses:

> *We hypothesise that the more tasks there are to do in a LearningBook, the number of tasks completed in the LearningBook will be lower.*
>
> *We hypothesise that the time taken to study a LearningBook is proportional to the number of tasks in the LearningBook.*

We found that there is a significant relationship between the number of tasks set and the number completed in a LearningBook. However, no significant correlation was found between the absolute number of tasks set in a LearningBook and the time spent interacting with a LearningBook. When we looked at the tasks completed as a percentage of all tasks set against the time spent in each LeaningBook we found a different picture which suggests that the more time a student spends, the more likely they are to give up before finishing all tasks.

More significant relationships were found when we examined the sessions that LearningBooks are composed.

As was true at the LearningBook level, there was a significant relationship between the percentage of tasks completed in each *session* and the number of tasks set.

We found that, on average for a session, the time taken to study a session is proportional to the number of tasks set which is a relationship that did not hold for LearningBooks.

Some of the results suggested that LBs12 and 20 were different to the other LearningBooks. These LBs were those over which students spent most time are distinguished by having only one session. This suggests that in future studies it might be worth looking at LBs with two or more sessions separately from LBs with a single session.

These results are based on the assumption that tasks are of the same difficulty in all sessions in all LearningBooks. The course team's view of task difficulty, measured as the average time expected per tasks ranges from 1.5 minutes for LB 12 to 5.6 minutes for LB 10. The student's experience was quite different. We adopted a similar measure of difficulty, the time spent per task completed, and found that students ranked the difficulty of LBs very differently from the course team.

A comparison of task difficulty with the number of tasks not completed showed a significant correlation.

These analyses seem to suggest the following student behaviour:

- students generally make a good attempt (greater than 80% on average) at all practical activities but seldom complete them all;
- students usually spend significantly less time on their practical work than expected by the course team;
- the more time students spend on practical work the more likely they are to stop before completing a significant proportion of the work;
- the students' perception of the difficulty of tasks is often different from that of the course team;
- there may be a significant difference between students' behaviour in relation to LearningBooks with multiple sessions from those with a single session;
- students normally tackle each session of a multi-session LB in a single sitting as recommended (although many take more sittings), but students tackle single-session LBs in multiple sittings (3 or 4);
- in general, a sitting usually extends for not more than about 30 to 40 minutes.

Our analyses lead us to the conclusion that looking at sessions is more revealing than LearningBooks as a whole, and there is a stronger relationship between LBs with multiple sessions, which suggests that student behaviour when there is only one session in a LearningBook is different from their behaviour when there are multiple sessions.

This data also suggests that the tasks completed tool gives a reasonable estimate of the actual number of tasks attempted. However, we know that the tool underestimates the number of tasks attempted so we shall be improving its accuracy and using it to perform similar analyses on a more substantial and representative collection of recordings.

Acknowledgements

Our thanks to Malcolm Macgregor for his diligence in implementing the Recorder and the Analysis tools and insightful comments on the research in general.

References

1. MacGregor, M, Thomas, P.G., Woodman, M., (1999) *Recording and Analysing User Actions in a Smalltalk Programming Environment.* TOOLS (USA).

2. Paine, C. & Thomas, P.G., (2000) How students learn to program: observations of study time behaviour. Research report: 2000/02. Computing Department : Open University.

3. Thomas, P.G., Martin, M., Macgregor, M., (1998) *Observing Students as they Learn.* ITiCSE 98, Dublin.

4. Thomas, P.G., (1998) *An Electronic Student Observatory.* Proceedings Frontiers in Education 98, p1120, IEEE. Phoenix: USA.

5 Smith, J.B., Smith, D.K. and Kupstas, E. Automated Protocol Analysis. Human-Computer Interaction, Vol. 8, pp 101 – 145, 1993.

6 Kivi, M.R., Grönfors, T. and Koponen, A. MOTHER: System for continuous capturing of display stream. Behaviour and Information Technology, Vol.17, No. 3, pp 152-154, 1998.

7 M206 LearningWorks © Open University

8 LearningWorks 0.7 © Adele Goldberg and Neometron, Inc and ParcPlace-Digitalk, Inc

9 M206 Computing: An Object-oriented Approach, The Open University (2000)

10 The Open University. http://www.open.ac.uk/

11 Computing Department. http://mcs.open.ac.uk/

12 Woodman, M. Griffiths, R., Macgregor, M., Holland, S., & Robinson, H. Exploiting Smalltalk Modules In A Customisable Programming Environment. Proceedings of ICSE 21, International Conference on Software Engineering. Los Angeles, 1999.

13 Thomas, P.G and Paine, C. How students learn to program: observations of practical work. Open University Computing Department report 2000/3, 2000.