RESEARCH AND EVALUATION DIVISION

www.ucles-red.cam.ac.uk

Quantifying support: Grading achievement with the Support Model

Ayesha Ahmed and Alastair Pollitt

University of Cambridge Local Examinations Syndicate

Paper presented at the IAEA Conference, Philadelphia, June 2004.

Contact details

Ayesha Ahmed, RED, UCLES, 1 Hills Road, Cambridge, CB1 2EU

ahmed.a@ucles.org.uk

UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

1 Hills Road, Cambridge, CB1 2EU

Quantifying support: Grading achievement with the Support Model

The Support Model

In our 2002 paper we described the support model for assessment (Ahmed & Pollitt, 2002). The essence of this approach is to assess students on the amount of support needed to complete a task. This differs fundamentally from the two principal current models which focus either on how well a student has performed on a task (judging) or on how much of a set of tasks a student has successfully completed (counting). For a discussion of the current models see Pollitt (1990). By aiming instead to measure how much help students need to produce an acceptable solution to a task we find we are not only measuring their level of achievement but doing so by assessing students in the process of learning, which means that we are getting closer to one of the aims of educational assessment that has always proved difficult to achieve – quantitatively assessing the process of learning.

The Question Answering Process

Our method for assessment using the support model is based on our psychological model of students' cognitive processing while they are answering examination questions (Pollitt & Ahmed, 1999). We consider the question answering process to consist of six phases. The first phase is the students' Learning, which happens before the exam in a traditional setting, and is what we are trying to measure in any examination. The next phase is Reading the question, and we find that it is during the reading phase that many misunderstandings and errors occur, which frequently prevent the students from showing us what they know or can do. We define a prerequisite for a valid question as that it must ensure that 'the students' minds are doing the things we want them to show us they can do'; this clearly cannot be happening if the students are unable to understand the question.

The next three phases of the question answering process are Searching, Matching and Generating. Searching is the spreading activation of concepts in the mind triggered by the reading of the question. Matching is the identification of relevant concepts, from which an idea of an answer is Generated. The final phase is writing which consists of turning this idea into, usually, a string of words.

These six phases were used as a framework for helping students towards an answer by prompting. For the purposes of this project the three central phases of searching, matching and generating were grouped together as one phase which we call *activation* (Anderson, 1983, Raaijmakers & Shiffrin, 1981). The prompting system is described in more detail below.

Research phase 1 – The Human Interviewer

This was a pilot phase reported two years ago (Ahmed & Pollitt, 2002) in which we carried out structured oral interviews with 15 and 16 year-old students. It was a simplified version of Phase 2 which will be described in detail in this paper.

Research phase 2 – The Human Computer

Oral interviews were carried out with 17 Geography students in which they were presented with a question and given support prompts until they reached a full answer. The students were all aged 15 or 16, and due to take their Geography GCSE (school leaving exam) in two month's time. The students were taking the Foundation tier exam, aimed at those who are expected to gain the lower grades (D, E, F or G). The questions that we used were from the Higher tier and were aimed at grades A*, A, B and C. The reason for this discrepancy was to ensure that the students did need the support prompts to answer the questions.

Interviews were recorded on audio tape, with two out of the seventeen also being recorded on video tape for presentation purposes.

One of the aims of Phase 2 was for the interviews to be more systematic than they were in Phase 1, so that the interviewer behaved more like a computer, since the eventual aim of the project is to computerise the whole process. This will be discussed in more detail below.

The Questions

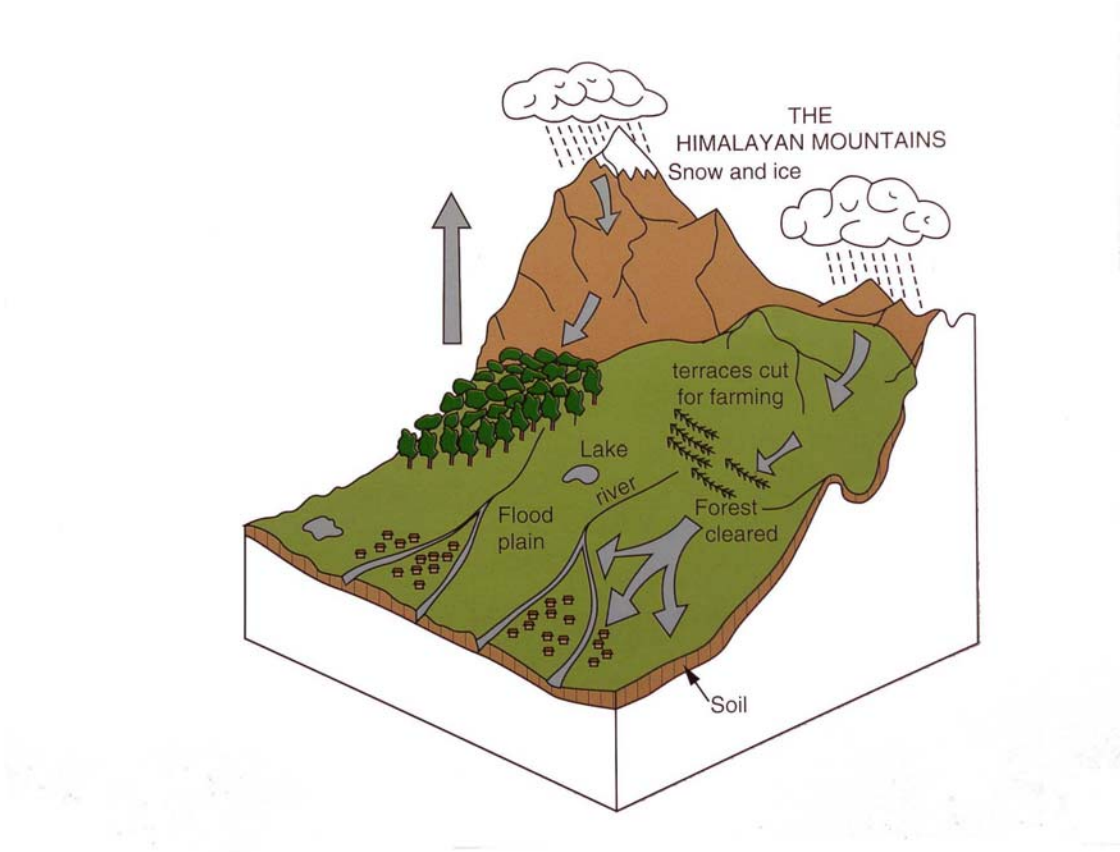
Two questions were used from past papers of the Avery Hill Geography GCSE syllabus aimed at 16 year-olds. The questions covered quite different topic areas but were both 'explain questions', asking students to 'Explain how' or 'Explain why'. The reason for this is that in subjects such as Geography and Science at GCSE these sorts of short answer questions are very common and a particular difficulty occurs with them. Students often do not know what level of explanation is required from them and they give too superficial an answer, or simply give a description. The examiners' expectations do not correspond to the students' expectations and the examiners tell us they would often like to be able to say to the students 'Yes, but why?'. The support model allows us to do this.

The two questions used in Phase 2 of the project are shown below. They were presented to the students in counterbalanced order.

Question 1

Changes to the hydrological cycle in the Himalayan Mountains

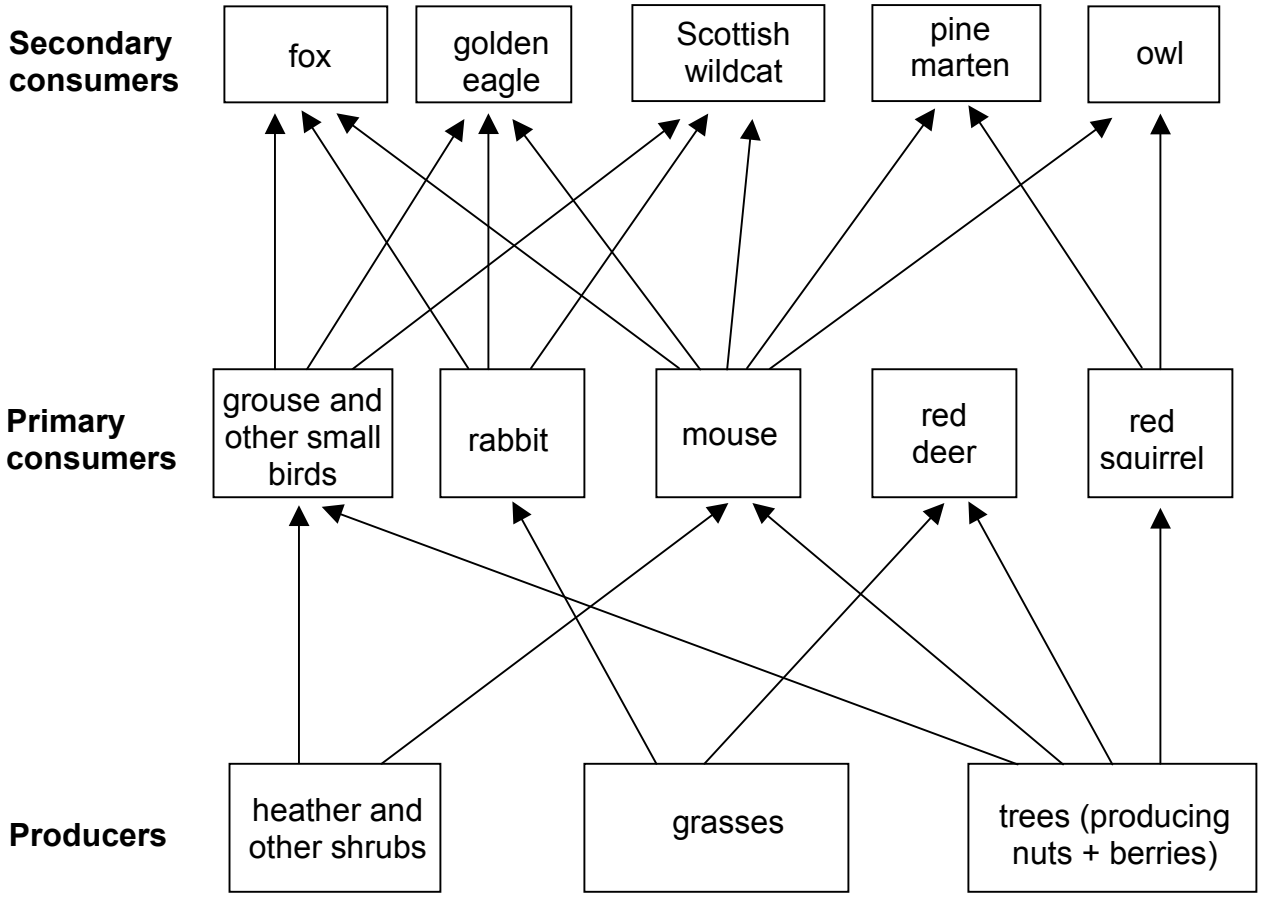
The foothills of the Himalayan Mountains are heavily forested. In recent years large areas of forest have been cut down for farming.



Cutting down trees affects the hydrological cycle. Explain how. [4]

Question 2

Study the diagram below. It shows some of the local species in a food web.



Which primary consumer does not have any natural predator? [1]

Explain why this might cause problems for the ecosystem. [4]

The Prompts

In our previous paper (Ahmed & Pollitt, 2002), we reported on the system of prompts based on our model of the question answering process. The first set of prompts correspond to the reading phase of the model. These are followed by concept activation prompts related to the central three phases of the model, and finally writing prompts. The sets of prompts are given strictly in this order, but not all prompts in a set are given each time as their need is determined from the student's response.

For each of the two questions a list of all the prompts we expected students might need was produced, with the help of an experienced Geography examiner. The prompts are listed in Appendix 1. The Reading prompts are devised by thinking through the possible difficulties students may have when reading the question: what do the words in the question mean? How might the question be misunderstood? Then Activation prompts

are developed by thinking about all the concepts that are relevant to the question and how we can prompt students' minds to activate them. These prompts also include asking for a deeper explanation. Finally the Writing prompts are based on the problems we expect the students to encounter when turning their mental representation of an answer into a text. They include prompting on how to structure an answer and to give a conclusion.

The result is that the prompts are an essential part of the task students face. All of them see the initial question and then have access to exactly the same prompts in the same order, simply missing out those that they do not need.

Our psychological model therefore gives us the general sequence of prompts and provides us with a script for prompting. The next consideration was how to order the prompts within the phases of the model. Within the reading phase some prompts relate to superficial aspects of reading the question whereas others are to do with a deeper understanding of the concepts in the question. Superficial reading prompts were given first, simply by repeating the question and, if still necessary, rephrasing the question. These were followed by prompts helping students to understand the concepts explicitly mentioned in the question, such as 'Do you know what is meant by the "hydrological cycle?" '.

The sequence of the activation prompts depends much more on specific features of the question, and on how the student begins to respond to the task. For example, in the water cycle question the activation prompts would usually start with the rain falling on the trees as this seems an obvious place to begin. However, they could start at a different point if the student responded at first with something about, for example, evaporation of rain from the leaves.

The order of the writing prompts is not so critical as they are concerned with structuring the answer once all of the relevant concepts have been activated. Obviously the ultimate prompt of 'giving the answer' would come last.

Also very important were the affective prompts. These were used throughout the process to give the students encouragement and praise. They also allowed the interviewer extra time to select the next prompt.

Sample Transcripts

Below are some excerpts from the transcripts to illustrate how the prompting worked and the ways in which some of the students responded.

Reading prompts:

The extract below illustrates how a superficial reading prompt helped the student from an initial incorrect answer to the right answer in the introductory part of Question 2.

A: Which primary consumer does not have any natural predator?

CA: Natural predator?

A: Yes

CA: ….. um … grass?

A: It's asking about the primary consumers

CA: Oh sorry. (long pause) deer, the red deer

A: The red deer, that's right

The following extract shows reading prompts being used in a more conceptual way, to help the student understand the question and build a mental representation of the task.

A: So here's the question and there's a diagram to go with it. It says; " Study the diagram. Cutting down the trees affects the hydrological cycle. Explain how." The diagram tells you; "The foothills to the Himalayan Mountains are heavily forested. In recent years large area of forest have been cut down for farming." So cutting down the trees affects the hydrological cycle. Explain how. Give me your first thoughts

T: (silence)

A: Anything you're thinking, just think it out loud for me

T: I think for uphill farming as well

A: Uphill farming?

T: Yes like sheep and stuff like that

A: OK. Have another look at the question. It's asking you about the hydrological cycle. Do you now what that is?

T: I think I've done it before but I can't really remember

A: OK. The hydrological cycle is the water cycle. Does that sound familiar?

T: (silence)

A: Do you remember learning about the water cycle?

T: I think we did it in like year 9

A: Year 9? OK don't worry. OK do you know what the water cycle is?

T: Yes when the rain comes and it goes and it evaporates back up

A: Perfect, that's it. You've got it. So what this is asking you is cutting down trees, how does that affect the water cycle?

T: Oh it can't condensate back up into clouds

A: Right OK. Let's have a look at the diagram a bit more closely then. Do you know what these arrows are showing you here?

T: Yes isn't that the way the rain's coming down

A: That's it, that's it exactly

Activation prompts:

This extract follows straight on from the above 'reading prompt' extract and shows how this student used the activation prompts.

A: So let's think what happens when there are trees there. What's going to happen to the rain? Where does the rain land?

T: (silence)

A: Do you know where the raindrops are going to land when there are trees there?

T: Is it on the um ….leaves isn't it?

A: That's it. Perfect. Yes, it lands on the leaves

T: Then it will evaporate up

A: That's right, some of it will evaporate up. What about if a lot of rain lands on the leaves? What will happen then?

T: Some of the rain will be absorbed by the plant itself

A: Right. If there's a lot of rainfall on the leaves. Do you think … what's going to happen to some of those?

T: It's going to be all weighed down, isn't it?

A: That's it. So what will happen when it's weighed down? With the raindrops. You're on the right track

T: It starts to drip or something like that

A: Perfect, that's it OK. It'll drip down. OK. So where is the rain going to land when it's dripped down?

T: Um … basically like near the roots so it's …

A: That's it. It going to land on the ground and ..?

T: It gets absorbed by the roots

Validity

We consider it a fundamental principle of validity that students should understand the question they are being asked to answer; after all, only then will their minds be doing the things we want them to demonstrate they can do. Can we really say we have examined a student if the question has failed to expose their learning to our view? An examination is a process that we carry out on students, it is not an object that students have to process. A metaphor makes the argument clear: we want to examine the kernel of their understanding, but we can't examine the nut if they can't crack the shell.

One of our criticisms of the current assessment models is that students who do not understand the question are very likely to score zero, and so be indistinguishable from others who understand the question but have no relevant knowledge. Our aim is to give credit proportional to the amount of knowledge each student has and we think the Support Model will come closer to achieving this.

As a consequence we believe that, in general, there should be no penalty for students who need support to understand the question. This does raise an issue, however. Some of the reading prompts relate to technical or semi-technical terms in the question that many teachers and examiners will feel students ought to understand if they have

learned the subject properly. The Support Model allows us to see how well students answer both before and after prompting on these terms, and we think that the question writers must decide whether or not to apply a penalty in each case.

Our examiners often ask a simple introductory sub-question, usually worth only one mark, before coming to the main 'Explain' question. Its function is often more to do with preparing the student for the main question, even though it is credited. This device was used in the Food Web question, and we chose to keep it in, although the examiner paid no attention to it in scoring. He did comment, however, that it had an effect similar to the reading prompts in that it "broke the ice", getting them past the initial phase of understanding what the question was about.

Note that technical term prompts in the reading phase are quite different from technical term prompts in the activation phase, where we are asking students to generate the terms for themselves. It is often possible for a student to give an adequate answer without using such terms, but examiners will always look for them and value a response more highly when they are used. Given this optional status of technical terms, is it right to penalise a student every time they don't use one without prompting? Our expert examiner decided, after discussion and long consideration, to limit the total penalty for this.

The heart of our procedure is prompting the student to activate the concepts that are important in answering the question; it is the lack of these that we aim to measure, since they indicate the amount of understanding the student has. The activation prompts are therefore the key prompts that, in any scoring system, the student should be penalised for needing.

The question of scoring the writing prompts is more difficult. If we wish only to assess the amount of understanding students have, then it makes sense to ignore the prompts they need to turn their understanding of an answer into a written answer to the question. Some, of course, would say that there is not much point in a student having knowledge if they are unable to express it, and so that the ability to write a coherent response should be credited too. Whatever view we adopt on this, however, there is a further argument: that the quality of a written answer will correlate with the quality of the student's understanding, and so makes a valid contribution to the assessment. Unfortunately, it also correlates with rather more general abilities, linguistic and cognitive, and we must be careful before we include these in the construct we are measuring. In the next phase of this research we will collect students' responses on computer and will then be able to explore the several options that this offers.

Grading

We envisage the main use of this assessment model to be formative, where teachers in classrooms could use the support model as a learning and assessment tool. If the assessment could be computerised it would provide an even more versatile tool.

However, we have also considered how support model assessment could be scored if it were to be used summatively.

The obvious way to think about scoring for the support model is to see each prompt as being worth a certain amount to the student. In this way we would have a negative

marking scheme with marks or grades being deducted when a student needed a prompt. It would be important that they needed the prompt, that is that their answer improved after the prompt was given. We would also imagine some prompts to be worth more than others, so that weighted marks would be deducted. In this way each prompt could be given a grade, for example a grade E prompt would be one that students who are 'worth' a grade E or below would need and others would not.

We worked on this idea with an examiner from GCSE Geography who was very experienced in marking students' work. In working together to analyse the transcripts we found that the prompts could not be said to have an intrinsic value as each was used differently by different students. For example, a single prompt might improve one student's answer from one that would score nothing using a traditional method to one that would score half marks. The same prompt could bring a different student's answer from half marks to full marks, or even from nothing to full marks. This meant that we could not assign a particular grade to each prompt.

Instead we decided to use a method which we call reversal. We reversed the order of the prompts and responses in each transcript so that the examiner first saw the final answer the student had given and then saw what help the student had needed to get there. In this way the examiner could focus on the prompts that had got the student to the full answer. It was also a more positive way of doing negative marking. He came up with a scoring system that he was happy with and that he felt gave each student the right grade and put the students in the right rank order.

He began by counting the number of activation prompts available to the students for each question. There were 22 activation prompts available to help with the concepts in the question, as well as some that asked for technical terms. He decided that a maximum of two prompts for technical terms was all that students should be penalised for in this question. This gave a total of 24 activation prompts. As described in the validity section above, we also decided that students should not be penalised for needing the reading or writing prompts, but only the activation ones.

The next step was to look at the actual transcripts, from which the examiner found that students never actually used more than 16 prompts. This gave us a notional maximum score of 16 for each question. One mark was then deducted for every activation prompt needed, and the final score was translated into a grade as follows:

A*	15/16
A	13/14
B	11/12
C	9/10
D	7/8
E	5/6
F	3/4
G	1/2

For example, a student who needed 10 activation prompts would have a score of 16 - 10 = 6. This would be equivalent to a high grade E. A student who needed only 4 activation prompts would have a score of 16 - 4 = 12, and therefore a low grade B. After grading

the transcripts in this way the examiner was happy that students had got the grade they deserved and that he had placed them in the correct rank order.

At this stage we have no better criterion for the validity of the system than that. The teachers of students like these predict the grades that they will get in the GCSE examination, and in August their actual examination results will be available, but both of these are measures of performance in an assessment based on reading and writing, skills they are not very good at, and in a setting that is daunting rather than supportive. If our 'grades' correlated well with those we would be rather disappointed.

To summarise, a system for grading has been developed which the examiner was happy with as a fair and valid system. Although not all prompts were worth the same to every student we ended up with a system where all activation prompts were treated as having the same weight – with one mark being deducted for every prompt used. This worked partly because we decided in advance that the key prompts were those dealing with conceptual understanding and the essential concepts in the question rather than those to do with reading the question or writing the answer. We were able to do this because of our systematic method for devising the prompts based on our model of the question answering process.

Students' Comments

The students gave us very positive comments about the support model process. Some even identified our key concern: the fact that it made them able to tackle questions on which they would have scored zero in a traditional system, just because they misunderstood the wording of the question.

> "I think it's quite a good idea because you then have the opportunity to get some help because in exams it's like one word that could stop a whole answer to a question and if you could get that one word then you might be able to get a really good grade instead of nothing."
>
> "You don't get all of the ideas, then when you start getting the ideas then more comes into your head."
>
> "….it's something really frustrating and stuff when you know and you just can't get it down."
>
> "It's kind of panicky when it's like … you've got to remember everything and if they were to give you some key words then you'd be like "ah, that's right!" It'd be a lot easier so people don't freak out and not know any of the question. It could be actually quite a good idea. I wish we could have it now."

The students clearly felt the benefits of an interactive oral exam. These issues will be different when it is computerised.

Confidence

The current assessment system in Britain fails those with low confidence by testing them in an isolated setting under stress. The support model is fairer to those lacking in confidence in two ways. First, students are encouraged to say something, however basic, to get them started on a question and hence to get some credit where in a traditional exam they would very probably just leave it blank and score zero. In addition,

the examiner is making a judgement based on the whole interaction with the student's mind and not just on the first attempt.

The following extract shows a student who didn't think she knew the answer when in fact she did. She has identified the 'problem' correctly.

A: "Explain why this might cause problems for the ecosystem"

V: What's the ecosystem?

A: That's OK, that's all right. The ecosystem is this system of plants and animals that are all connected together

V: Well I wouldn't be able to think of a problem but the red deer wouldn't die out

The next extract shows a student who could not think of a technical term but when he is told it he says he did know. We have no way of being certain that he is telling the truth here but the students appeared to be being honest and it was not a situation in which they would have felt they needed to say they knew when they did not.

A: And do you know what it's called when the rain lands on the leaves instead of falling straight to the ground?

F: No

A: Interception

F: Oh yeah I know it now you've said it

The extract below reinforces our feeling that the students were being honest about what they did and did not know.

A: Right. And do you know what it's called when the rain lands on the leaves, instead of falling straight to the ground?

G: No

A: Interception

G: No, I didn't know that

The issue of confidence and its effect on performance will be very different when the assessment is carried out on computer rather than as one-to-one interviews.

Discussion and Conclusions

In this final section we will begin by discussing the next two phases that we see in the research programme to develop useful tools using the support model.

Research phase 3 – The 'Human' Computer

The next phase of this research will take us one step closer to the ultimate goal of administering and scoring this system on computer. We intend to present the questions and their prompts on screen. We will require the student to make an initial response into a text box, and then to amend this response after each prompt, until it is good enough for full credit.

The prompts will still be selected by a human, based on reading the student's current answer. To begin with, the researcher will sit with the student throughout the process to support them with encouragement and praise, but we intend to try a version in which the researcher will be out of sight and not interacting directly with the student at all. It is possible that when there is no adult observing what the students are writing, they will write differently, perhaps by being more likely to try an answer they are not sure of. Because the following phase will be fully automatic we want to see the range of responses that it will elicit in order to design the software.

We will need to decide how to handle a null response. It is likely that some students will not type anything when given just the first prompt, because the first prompt will actually be just the original question. How long should we wait before giving the next prompt, which will be the first real support prompt? Once students are used to this form of computer interaction we do not think it likely that they will continue to give no response, but the system will need a defined procedure, for safety.

During this phase we will explore ways of giving more control over the prompting system to the student by, for example, making some of the prompts available to them as clickable hot links or in drop-down menus. Such ideas must be approached with care, however, since we do not know how students' willingness to use prompts of this kind will depend on their confidence and willingness to take risks.

Research phase 4 – The Intelligent Computer?

In its most primitive form the Support Model formally consists of a sequence of prompts. A student is invited to modify their initial response after each and every prompt in the sequence. But computers are capable of some intelligence. We hope to develop a system which will be able to evaluate the response after each prompt, and decide whether or not each subsequent prompt is still needed. This would avoid the tedious presentation of every prompt to a student and the irritation that would doubtless induce. In the formal presentation below , 'State' indicates the current state of a vector internal to the programme which indicates which prompts may still need to be presented.

start

State : < 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 >

Prompt 1: [ie, the question]

State : < 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 >

Response 1: 'Dunno {a very poor answer}

---Evaluation---

State : < 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 >

Prompt 2: [This is what the question means]

State : < 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 >

Response 2: 'Oh! In that case, yyyyyyyyyyyyyyyyyyyyyyyyyyy' {a better answer}

---Evaluation---

State: < 0 0 0 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 >

Prompt 4: [xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx]

. . . . . .

Response n: 'yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy' {a full credit answer}

---Evaluation---

State: < 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 >

finished

In this example, the student is given the 'first' prompt (the question) and does not reply with a response worth anything. The second prompt then explains the meaning of the question more fully, or by using different words, and the student is then able to make an attempt at a response. As a result of this response, the 'evaluation' is that only one of the next six prompts is needed. Prompt 3 is not given (and nor will be prompts 5, 6, 7 and 8) but prompt 4 is given. Following the response to this, prompt 9 may be given next. The process continues until the 'evaluation' has decided that no more prompts are required; this is the stopping criterion for the procedure.

'Evaluation' as used above is the intelligence we want programmed. Computer software has long been used for marking objective questions, and for several years has been applied to rating essays. Recently, some success has been reported in computer assessment of the more difficult short answer question of the kind we are using here (Mitchell, Russell, Broomhead & Aldridge, 2002; Sukkarieh, Pulman & Raikes, 2003). A program being developed for us by Pulman and Sukkarieh may have, or be adapted to have, the capability we need – to recognise words or phrases in a response that indicate that a certain prompt is unnecessary.

On this basis our stopping criterion would be that every prompt would be judged unnecessary. There is, though, an alternative criterion that might be as or more effective. An essay rating technique that is able to evaluate the content students write, such as Latent Semantic Analysis (Landauer & Dumais, 1997; Burstein, Kaplan, Wolff & Liu, 1996) should be able to judge when an answer is good enough to merit full marks, and this would amount to another way of deciding that no further prompts are needed.

If we use the original criterion, of a full vector of zeros, there is still a further possibility. Automarking software could evaluate the quality of the final answer, giving a second score in addition to the score derived from the prompts needed. This quality score would presumably measure the student's ability to write a coherent explanation, given the necessary concepts, and would therefore complement the 'support' measure of the student's understanding of the concepts.

Generalising the model

So far we have only tried this approach with 'explain' questions. In contrast, earlier researchers have tried interactive testing techniques with problem solving tasks (as far as we can see always in mathematics, eg McAlpine & Ware, 2003; van der Bergh, undated). The principle difference we have seen between our tasks and these is the distinction between the mostly subconscious processes of searching, matching and

generating described in our theory of the question answering process and the far more explicit conscious thinking involved in intentional problem solving. Recently, however, Shikata (2002) described high level mathematical problem solving with a model based metaphorically on warfare whose four phases – Conceptual, Strategic, Tactical and Battle – are very similar to our Reading, Searching, Matching and Generating/Writing. We have always understood that in some tasks, notably essay writing, it is appropriate to combine the last two phases of our model in the way that Shikata does, and the similarities between his model and ours leads us to think it probable that a system of support prompts could be systematically constructed for mathematical problem solving too.

Setting up the prompts for a support model question will always require a significant investment. Even in the case of the 'explain' questions our underlying theory can only indicate the nature and sequence of the prompts that might be needed. To extend the method to other sorts of questions, whether problem solving or some other kind, we need to understand how students think while trying to solve them. To do this we advocate using a video protocol method (Ahmed & Pollitt, 2000; Lyle, 2003). Students are recorded in the process of answering questions, and the recording is played back to them immediately afterwards to stimulate rehearsal of their thinking processes. At any point the tape can be paused to allow the students to report at length on the concepts that were active in their minds at that point. Although experience with new question types will doubtless allow us to shorten the process of generating the set of prompts we feel that this is one reason why the support model may be more appropriately applied in formative assessment, where questions can be re-used many times to repay the investment.

Reliability and Validity

There may seem to be a disadvantage associated with the Support Model in that it will lead to students scoring partial credit on some questions on which they would otherwise have scored zero. This will necessarily reduce the total score variance, and hence is likely to reduce the value of any reliability index calculated. But consider what happens with the scoring system we have proposed: in an ideal test a given student should score the same, eg 'D', on every question. If, as with current exams, they instead score 'D' on half and zero on the others, because they didn't understand the questions, then their average score will be 'F' (since on a scale that runs from A to G a zero is equivalent to 'H'). This is manifestly unfair. If moving to the support model loses some of the score variance, it is therefore clear that all of the score variance we lose is invalid. It is likely that we would observe the effect of this as an increase in the inter-item correlations in the exam.

Because we are uncovering learning that has hitherto not been credited we would hope that there is a net increase in validity. Students being assessed in this way will spend more of their time with their minds 'on task', doing the things that we want them to show us that they can do.

References

Ahmed, A & Pollitt, A (2002) *The Support Model for Interactive Assessment*. International Association for Educational Assessment Conference, Hong Kong.

Ahmed, A & Pollitt, A (2000). *Observing Context in Action*. International Association for Educational Assessment Conference, Jerusalem.

Anderson, JR (1983). *The architecture of cognition*. Cambridge MA: Harvard University Press.

Burstein, J, Kaplan, R, Wolff, S & Liu, C (1996). *Using lexical semantic techniques to classify free-responses*. In Proceedings from the SIGLEX 1996 workshop; Annual meeting of the Association of Computational Linguistics, University of California, Santa Cruz. http://www.ets.org/research/erater.html.

Landauer, TK & Dumais, ST (1997). A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 25, pp 259-284.

Lyle, J. Stimulated recall: a report on its use in naturalistic research. British Educational Research Journal, 29, pp. 861 – 878.

McAlpine, M & Ware, W. (2003). *Introducing computer-assisted assessment in Scotland: Laying the foundations for an integrated approach*. International Association for Educational Assessment Conference, Manchester.

Mitchell, T, Russell, T, Broomhead, P & Aldridge, N (2002). *Towards Robust Computerised Marking of Free-Text Responses*. The 6th International Computer Assisted Assessment (CAA) Conference, Loughborough University.

Pollitt, A & Ahmed A. (1999). *A New Model of the Question Answering Process*. International Association for Educational Assessment Conference, Slovenia, May 1999.

Pollitt, A (1990). Giving students a sporting chance: assessment by counting and by judging. In Alderson, C. & North, B. *Language Testing in the 1990s*, pp 46-59. Macmillan: London.

Raaijmakers, J G W & Shiffrin, R M (1981). Search of associative memory. *Psychological Review,* **88**, 93-134.

Shikata, Y (2002). *Strategy and tactics in solving mathematical problem*. International Symposium on Test Development for University Entrance, NCUEE, Tokyo.

Sukkarieh, JZ, Pulman, SG & Raikes, N (2003). *Auto-marking: using computational linguistics to score short, free text responses*. International Association for Educational Assessment Conference, Manchester.

van der Bergh, N.(undated). *AIM documentation.* University of Ghent. http://allserv.rug.ac.be:80/~nvdbergh/aim/docs/

Appendix 1 – The Prompts

Question 1

Reading Prompts

Repeat question

Re-phrase question

When people cut down trees, how does that affect the water cycle?

Help them understand the question

Do you understand what is meant by the hydrological cycle?
The hydrological cycle is the water cycle.
It is the way water moves from place to place.
The diagram shows you how water moves from place to place.

Help them understand the diagram

Do you know what the arrows are showing you?
They are showing you the way the water moves from place to place.

Activation Prompts

If no real first answer:

Think about what will happen when there are trees there.
What will happen to the rain?
This part of the diagram shows you what happens when there are trees.
This part of the diagram shows you what happens when there are no trees.
Concept: Interception
Do you know what happens to rain when it falls on trees?
Where will the raindrops land?
Which part of the tree?
Do you know what it is called when the rain lands on the leaves instead of falling straight to the ground?
Have you heard of Interception?

Concept: Infiltration

Do you know what will happen if a lot of rain falls on the leaves?
Do you know what will happen to the rain that lands on the ground?
Do you know what the soil will be like under the trees?
Think about the leaf litter on the ground.
Think about whether or not the soil will be able to absorb water.
Do you know what it is called when the rain goes down into the soil?
Have you heard of Infiltration?

Concept: Transpiration

Do you know what happens to the water that is in the ground?

Think about the tree roots.
Where does the water go when the tree roots have taken it up?
Do you know what it is called when the water goes up the tree roots?
Have you heard of Transpiration?
Concept: Evaporation
Then what happens to the water when it is on the leaves?
How does the water get back into the atmosphere?
Do you know what this arrow is showing?
Do you know what it is called when the water goes back up into the atmosphere?
Have you heard of Evaporation?
Concept: Surface run-off
Do you know what will happen to the rain if there are no trees?
Do you know how the soil will be different if there are no trees?
What happens to the water on the ground if it can't all be absorbed into the soil?
What will happen to the rain that is resting on the surface?
Do you know what this arrow is showing?
Do you know what it is called when no more water can be absorbed and it flows along the ground?
Have you heard of Surface run-off?
What will happen if there is a lot of surface run-off?
Asking for a more specific answer
And can you explain how that happens?
And can you explain why that happens?

Writing Prompts

Can you have a go at a final answer now? You don't need to write it down – just tell me what you're final answer to this question would be.

Asking for a conclusion

So, how would you conclude that answer?
What is the final effect of cutting down the trees?

Structuring the answer for the task

Look back at the question….
….it is asking you for what happens when there are no trees.
….it is asking about the effects on the hydrological cycle.
…. it is asking you for an explanation

Clarification of their answer

Giving them the answer

Affective Prompts

Encouragement all the way through :

Well done, you're doing really well ….

To introduce prompts without de-motivating them:

You're on the right track now – how about ……

That's good, now think about it this way ….

If lack of confidence at the beginning:

Don't worry we'll work on this together – it's meant to be hard – I'll help you work out the answer ….

Question 2

Here is a question and a diagram that goes with it.

This diagram shows a food web.

There are Producers, Primary consumers and Secondary consumers.

First question part:

Try to say all your thoughts out loud.

Look at the primary consumers.

Do you know what these arrows mean? (from primary to secondary)

Main question:

Look at the next part.

Can you try to give me your first idea of an answer? Try to say all your thoughts out loud.

Reading Prompts:

Repeat question

Help them understand the question

Do you know what is meant by ecosystem?

The diagram shows you the ecosystem: it is the system of plants and animals.

Help them understand the diagram

Do you know what is meant by producer? Primary consumer? Secondary consumer?

Do you know what these arrows mean?

Re-phrase question

If the red deer doesn't have any predators what might this mean for the ecosystem?

How will the ecosystem change if the red deer has no predators?

Activation Prompts:

Asking for an explanation/giving a concept

What would happen to the red deer if no other animals were eating them?
How would this affect some of the other plants and animals in the diagram?

Producers

How would this affect the grasses and trees?
How would this affect the amount of grasses and trees?
Look at the arrows that are going to the red deer.
Think about food supplies for the red deer.

Primary consumers

What will happen to the rabbits?
Look at the arrow going to the rabbit.
Think about the numbers of rabbits.
Think about the food supplies for the rabbits.
When the rabbits and the red deer both want to eat the same food what do you call this?
This is what we call **competition**.

What will happen to the red squirrels?
Look at the arrow going to red squirrel.
Think about the numbers of red squirrels.
Think about the food supplies for the red squirrels.
When the red squirrels and the red deer both want to eat the same food what do you call this?
This is what we call **co**m**petition**.

Secondary consumers

Now look at the arrows going up from the rabbit/red squirrel
What will happen to the owls/foxes/pine martens ….?
Think about the numbers of owls/foxes/pine martens ….
Think about the food supplies for the owls/foxes/pine martens ….

If going very well ….

What about the mice and the grouse?
Are they affected differently from the rabbit and red squirrel?

Asking for a more specific answer

And can you explain how that happens?
And can you explain why that happens?

Writing Prompts

Can you have a go at a final answer now? You don't need to write it down – just tell me what you're final answer to this question would be.

Asking for a conclusion

So, how would you conclude that answer?

How would this affect the whole ecosystem?

Structuring the answer for the task

Look back at the question….

…. it is asking you for an explanation

…. it is asking you about problems for the ecosystem.

Clarification of their answer

Giving them the answer

Acknowledgements

The authors would like to acknowledge the considerable help they have themselves been given by two colleagues: Hannah Fisher-Hoch and Jane Fidler were model supporters in carrying out some of the interviews and processing the results. We would also like to thank Robert Walker, the senior examiner in Geography, whose experience and judgement we learned to trust throughout the work.