Assessment in Education, Vol. 10, No. 3, November 2003

Carfax Publishing
Taylor & Francis Group

Using Computers to Assess New Educational Goals

JIM RIDGWAY & SEAN MCCUSKER
School of Education, University of Durham, Durham, UK

ABSTRACT *There is a remarkable consensus worldwide on 'new' educational goals. These emphasise problem solving using mathematics and science, supported by an increased use of information technology. Change can be difficult: first is the problem of communicating new goals; second is their alignment with old assessment systems. Well-designed assessment can solve both these problems. Here, computer-based tasks are described which exemplify new goals, and could be used to promote desirable educational change. Computers can make a unique contribution to assessment. They can present new sorts of tasks, where dynamic displays show changes in several variables over time. Interaction makes computers well suited to the assessment of process skills—discovering rules, finding relationships, developing effective strategies—by the use of simulations, microworlds and interactive games. Students can work with complex realistic data sets, using professional methods. The paper illustrates these claims, and describes student strengths and weaknesses observed on live tests.*

Changing Educational Goals

An interesting development worldwide has been the emergence of a shared rhetoric on education, and the emergence of an overlapping set of educational goals, independently, in different countries. These can be observed in the similarities of statements related to National Curricula or their equivalents, and the statements of educational aims from government ministers around the world. Examples include Japan, Hong Kong, Singapore, Hungary, New Zealand, and England (see, for example, http://www.minedu.govt.nz; http://www1.moe.edu.sg/iteducation/). In countries where the control of education is less centralised, such as the USA, analogous statements can be found in influential documents prepared by professional organisations such as the National Council of Teachers of Mathematics (2000), and in the Technology Standards developed jointly by the International Society for Technology in Education, and the National Council for Accreditation of Teacher Education, which are having a major impact on State policies (see Taylor, 1996); the same phenomenon can be seen in other federal systems (e.g. Queensland, Australia).

These curriculum ambitions can be summarised briefly. There needs to be an increased emphasis on mathematics and science, supported by the increased use of

ISSN 0969-594X print; ISSN 1465-329X online/03/030000-00 ©2003 Taylor & Francis Ltd
DOI: 10.1080/0969594032000148163

information technology (IT). Within these subjects, there will be an increased emphasis on higher-order thinking such as problem solving and communication skills, with a focus on solving practical problems.

Change in educational practice is often found to be difficult (Fullan, 1991). The sorts of changes necessitated by the new political agenda for education are particularly difficult to bring about for two distinct reasons. First is the problem of communicating the new vision to be attained; second is the inherent stability of educational systems, in the form of elements which lock them in place, such as rigid assessment systems, or inflexible programmes for professional development. There are major problems associated with communicating new ideas. All constructivist accounts argue that one sees the world through the lenses of existing knowledge structures. Communication about novel practices therefore runs real risks that genuine novelty will be 'adapted' (in the Piagetian sense) to existing knowledge structures. We believe that vivid exemplification of new goals is the way to tackle this problem. Students and teachers need to work with tasks (and their associated mark schemes) in order to understand the new cognitive goals to be reached. This exemplification is a key objective of the work reported below.

A second major barrier to educational change can be the high-stakes assessment systems which are in place. Successful innovation depends critically on the alignment of the key elements within a system—curriculum goals, curriculum implementation, professional development, and the assessment system (see Webb, 1999). Here, we focus on the alignment of ambition and assessment. Assessment systems have a profound influence on professional practice (Clarke & Peter, 1993). High stakes assessment drives educational practice. A key activity for the assessment community must be to design measures of attainment which relate to new educational goals. Without them, new educational ambitions are likely to be ignored. The goals described above, common to a wide range of countries, are unlikely to be achieved unless there are ways of measuring performance on these goals. The development of assessments aligned to new educational goals is the second key objective of the work reported below.

Learning to Use New Tools

From a sociological perspective, it is clear that IT is at the centre of a cultural vortex which is bringing about radical social change in terms of patterns of employment and methods of communication; at the level of intellectual discipline, IT has had a profound effect on the cultural practices associated with every academic discipline over the last 20 years. For example, imagine anyone doing research in literature who did not use a word processor; a statistician who did all calculations by hand; a physicist who logged their laboratory data after direct observation. Such people may still exist, but are the Zeppelins of their profession. The profound changes associated with IT use underpin the political will worldwide to reform curricula—and also explain the emergence of the common goals across countries described earlier. To respond adaptively in fast-changing fields, it is necessary to understand something of the complexity of the cognitive and cultural changes that are taking place. Vygotsky's

work (e.g. Vygotsky, 1981) on the relationships between culture, intellectual tools, and cognitive development provides a useful starting point. In the context of education, Vygotsky and people working in the Vygotskian tradition (e.g. Davydov, 1995) argue that education should focus on developing those intellectual tools which give individuals most power over their environment, and that students should be made aware of the usefulness of the tools they are acquiring, in order that they see the reason to invest intellectual effort in the acquisition of these tools (see Ridgway, 1997a, b).

Vygotsky has a number of important ideas for us:

- culturally important tools—here, computers—have a role of defining cultures;
- computer-based tools will change: thinking, the ways that individuals see the world, and their ability to function effectively in the world;
- learning can begin in piecemeal ways, but fluent use of computer-based tools needs to be integrated with other intellectual skills before it constitutes a set of higher-order skills;
- students should be exposed early to the power that intellectual tools can give them in different situations, in order to motivate learning;
- there may well be differences in the minds of people raised in pre-computer cultures (e.g. many teachers) and people raised in computer cultures (e.g. students), in just the same way that people raised in different language cultures can see the world in rather different ways.

The Value of Cognitive Abilities is Culturally Defined

Cognitive abilities that are highly valued in one culture may have little currency in another, where they have been rendered redundant by a new technology. Consider pre-literate societies. Rote memory for large volumes of material was valued highly, for good reasons, such as the need to maintain a consistent version of the law. In tenth century Iceland, the law-speaker recited a portion of the law at each assembly of their parliament. Similarly, orators in Greece and Rome developed methods to remember speeches; other ancient thinkers developed methods for remembering people's positions at a meeting, and lists of objects in order, which are still given as examples of 'metamemory'. These methods include the method of loci (mentally place objects along a familiar route); learn a rhyme and associate objects with the rhyme ('one is a bun, two is a shoe …'). These mnemonics are interesting historically, but have rather little use today—perhaps with the exceptions of remembering facts in a particular order for examinations or for party tricks (see Baddeley, 1990). In literate societies, many of these devices are largely unnecessary, because information can be written down and accessed when required. The big idea we share with earlier civilisations is that problems of knowledge acquisition and retention should be studied, and not left to chance—but the sorts of knowledge to be acquired retained and used have shifted as function of the technologies we use.

The idea that one should be aware of cognitive processes in general, and one's own knowledge structures in particular has been the focus of a good deal of research

(much of it triggered by Flavell's (1976) early work on metamemory). As well as an increasing literature on 'metacognition', and the importance of developing these skills as part of learning the current curriculum, educators face a new challenge in that new technologies make new demands on cognition, and require the development of new sorts of thinking skills. Consider the use of spreadsheets. Spreadsheets facilitate fast calculation and recalculation in situations where a number of factors interact, and are excellent for exploring 'what if' situations. The essence of this activity is to change the assumptions made initially, and to explore the implications. Even 20 years ago, such intellectual explorations would have been considered as indications of very sophisticated thinking amongst adult financial planners, and people modelling complex situations; now we expect children to do it. Consider also the processes of literary criticism at undergraduate level. Students are asked to comment on the messages the authors wish to communicate; the impressions they wish to convey to their audience; the information they present (and choose not to present); and the rhetorical devices they use. In a pre-computer era, these would seem rather difficult intellectual tasks with which to engage children. However, in the context of accessing web pages, primary school children can be asked questions about authors' communicative intentions, and their uses of rhetorical devices. The web provides a relevant context and powerful reasons why these questions are important (e.g. Downes & Zammit, 2000).

The relationship between technology and the development of cognitive skills is complex. New applications provide an impetus for the development of new cognitive skills—but as we get smarter, we develop new applications, which create further pressures for the development of new cognitive skills. In short, technology and cognition co-evolve at both the individual and social level. Education might be an important element in this process (as politicians believe, worldwide), but again the relationship between the elements is complex. Technology provides a set of pressures for educational change in the direction of developing new intellectual skills. Technology also provides a unique medium for assessment. As we learn more, we are able to create new applications, describe new sorts of intellectual skill, and develop new sorts of assessments which access these new skills.

The work described in this paper focuses on the development of new sorts of assessment tasks in the light of new cognitive demands, and the potential of technology to support new sorts of assessment.

Mapping a New Cognitive Agenda

The previous section argued that the definition of what is worth knowing is being challenged by emerging technologies. A number of components of the new educational agenda can be mapped out. These include: the promotion of meta-knowledge; using new representations and symbol systems; and modelling complex processes and problems. A desire to develop higher-order thinking can be traced back at least to Vygotsky; there is an honourable tradition of assessing problem solving via the use of extended tasks, such as those developed by the APU (e.g. Archenhold *et al.*, 1988). However, the computer offers some unique features in terms of representa-

tion, interaction, and its support for modelling. Some of the cognitive skills that are now receiving greater emphasis are described below; some exemplar tasks are shown in Section 6. Other materials have been developed by the OECD Programme for International Student Assessment (PISA) (http://www.pisa.oecd.org/).

Promoting Meta-knowledge

As the number of sources of information expands dramatically, there is a need for students to develop sophisticated theories-in-action about knowledge. These theories should include accounts of the nature of knowledge—its generation, and the various functions it serves (including its use as just another rhetorical device!). Students also need to know about their own knowing—what they do and do not know, how they acquire, lose, and change their own knowledge—and how they control their cognitive processes when solving problems. Here we will focus on the development of metacognitive skills. Developing these skills may well be difficult; studies of expert-novice differences (e.g. Anderson, 1993) suggest that the time-line for the development of metacognitive knowledge is likely to be decades rather than years. However, this evidence is based on earlier, pre-IT-based learning activities in schools; it may well be the case that metacognitive knowledge can be developed far more quickly, once it is taken seriously as an educational goal (e.g. Adey & Shayer, 1994; Schoenfeld, 1985). Most tasks which are unfamiliar to students, and which require extended chains of reasoning call upon these skills. Students have to plan their activities, work systematically, know what they know and do not know, and must monitor their progress, and respond appropriately. Working on unfamiliar tasks with appropriate coaching can lead to both better performance and to the development of better metacognitive skills (Schoenfeld, 1985). Most of the tasks described below, which were developed for use in The World Class Arena (www.worldclassarena.org) require the use of metacognitive skills.

Using New Representations and Symbol Systems

An important intellectual skill is the ability to use a variety of representations when solving problems. Examples include the ability to represent events shown in a video in graphical form, to express a verbal description of some situation in terms of algebra, and generally to be fluent in translating between representations such as words, diagrams, algebra, and graphs. A number of authors (e.g. Schubert, 2000) argue for the development of the ability to work with new representations and symbol systems. Existing evidence from international surveys (e.g. TIMSS: http://www.csteep.bc.edu/timss) shows that students find it very difficult to solve problems which require abstract representations. Computer presentations allow a variety of representations and symbol systems and allow easy translation between them—indeed information and computer technology (ICT) is a major source of new representational methods. The new educational agenda must address the needs of students to engage with representations they have not encountered before. *Speed*

(Figure 1), and *Cowpats* (Figure 3), provide illustrations of the use of new representations and symbol systems.

Finding rules and relationships. Effective problem solvers 'know what to do when they don't know what to do'. Presenting problems that require students to find rules and relationships via careful exploration is easy to do via computer-based microworlds, and is far harder using other media. *Sunflower* (Figure 2) and *Pyramids* (Figure 7) provide illustrations.

Constrained decision making. A common task in many areas of design is to make decisions among alternatives which are constrained in some way—for example, the choice of material for some component of a bicycle will be a function of cost, weight, strength, and resistance to wear. Tasks of this sort are easy to create, via databases. It is common for a number of materials to satisfy the minimum requirements to perform a particular task. The designer must then evaluate the alternatives against the original criteria, and secondary criteria (e.g. colour) that might emerge, later. *Holidays* (Figures 8a to 8d), and *Fishing* (Figures 9a to 9c), provide illustrations of constrained decision making.

Handling complex data. Before the widespread use of computers, handling large and complex data sets was extremely difficult to do—imagine, for example the effort that would be required to create the statistical analyses of test performance done routinely as part of any examination process. The skills of working with complex data sets must be part of the new educational agenda. With appropriately designed interfaces, the computer can present tasks based on complex data sets in a way that students find easy to use. *Fishing* (Figures 9a to 9c), and *Holidays* (Figures 8a to 8d), provide illustrations of handling complex data.

Modelling complex processes and problems. This challenge takes students to the forefront of academic work. Almost by definition, developing the skills to model complex processes such as those associated with health, crime or our environment are likely to be difficult. Computer-mediated tasks can be very helpful. Complex data can be presented in ways students find accessible. Conventional wisdom suggests that 9-year-olds can only handle problems with one independent variable. Our evidence shows that they can do far more than this. We attribute the enhanced performance to the fact that the contexts are familiar, and dynamic displays enable students to engage with the core intellectual task without having to struggle with the computational aspects. *Chasm* (Figure 5), and *Cowpats* (Figure 3), provide illustrations of modelling complex processes and problems.

Developing Computer-based Assessment

Assessing Problem Solving

One can conceive of the problem solving process as a cycle, which requires the solver

to: understand and represent the problem in some way, and hopefully in multiple ways; to explore these representations and create some solutions; to test solutions against the original specification; and (after several iterations) to explain why a particular solution or solutions do actually answer the problem as posed.

When assessing problem solving, it is unlikely that any single problem will address each of these phases; over a whole test, however, there should be good coverage of each phase. Some examples from the World Class Arena (WCA) are presented below, to show how this can be done.

Understanding and Representing Problems

Traditional educational goals such as the ability to interpret tables and graphs, and to translate information coded in one representation into information coded in another representation continue to be vital skills for mathematical and scientific literacy. Computers allow fast and reversible transformations of information from one representation to another, thus revealing the relationships between them. Computers also provide ease of access to a variety of presentation, such as video, as well as new representations and symbol systems. The ability to learn to use new systems is part of the new educational agenda.

We are exploring a variety of tools for students to use when solving problems—such as graph drawing tools which allow students to sketch graphs which capture the qualitative aspects of realistic situations, without the need either for plotting points, or the use of algebra-based graph plotters.

Speed (Figure 1) is a task for 13-year-olds which presents students with a video of a car travelling along a road. They are required to represent the journey as a graph, which they build from line segments. Students find this task easy.

Problem Exploration

The computer can facilitate the creation of microworlds for students to explore, in order to discover hidden rules or relationships. Examples are simulations of laboratories or realistic settings, where students conduct experiments in order to discover rules and regularities, and games, where they set out to develop winning methods. The desire to assess process skills in science is not new. Traditionally, students would be presented with tasks in laboratories, or would be required to keep logs and portfolios of their laboratory work. These approaches suffer a number of drawbacks. The laboratory setting can introduce elements which reduce the reliability of the assessment, such as instruments which fail to function properly, or materials whose properties are less than ideal. Students are required to manipulate apparatus physically—chance differences between students in terms of their previous exposure to particular equipment can both reduce reliability, and add an extra cognitive load to the intellectual task being performed. Some education systems are unwilling to accept teacher ratings of students for the purposes of high stakes testing, with the result that process skills in science are not assessed at all. Computer-based assess-

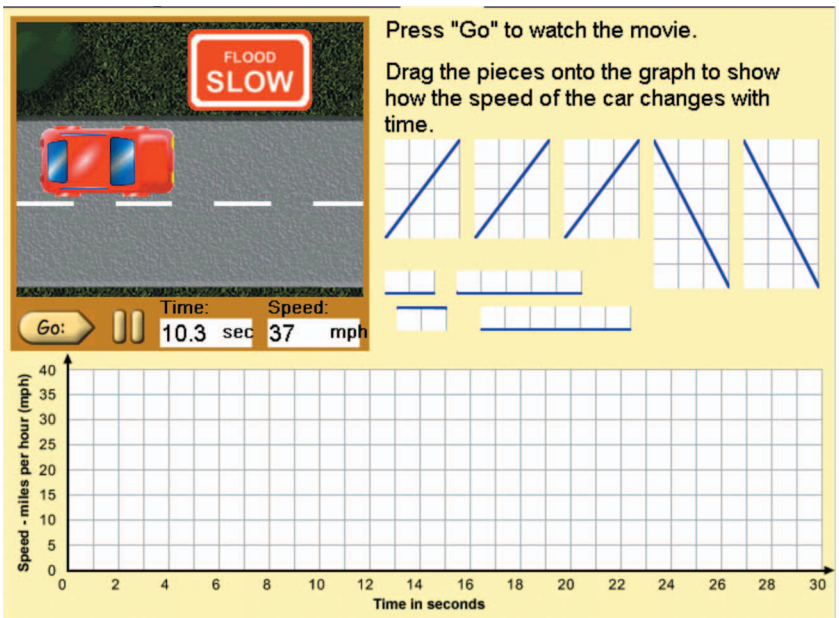

FIG. 1. Speed.

ment permits the assessment of these valuable aspects of learning science, at modest cost. Computer based assessment could also be used to moderate teacher grades.

A range of different process skills can be identified, which include:

- working systematically—for example, choosing tests systematically and recording results systematically (WCA tests show that students are often poor at this);
- controlling variables.

Sunflower (Figure 2) is a task for both 9- and 13-year-old students which presents students with the task of growing the world's tallest sunflower, using some combination of the two nutrients that are provided. The task rewards systematic work such as controlling variables, recording results, and careful exploration of the search space.

Creating Solutions

The creation of solutions lies at the heart of problem solving. The choice of the most appropriate process is a function of the task; however, a number of general purpose strategies can be described. Here, we consider generating and testing hypotheses, discovering rules, and finding relations.

Generating and Testing Hypotheses

Cowpats (Figure 3) provides information on the size of a cowpat and the presence of flies, worms, and beetles over a period of 100 hours. Students make deductions

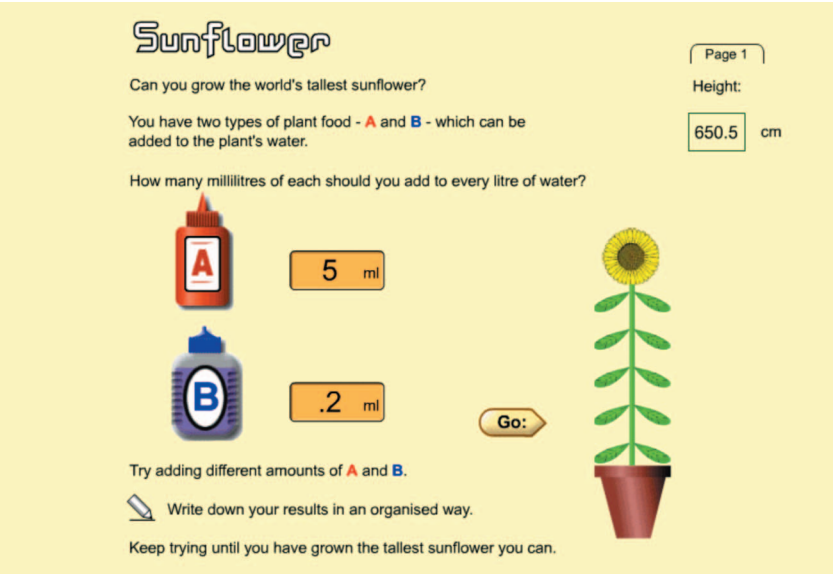

FIG. 2. Sunflower.

about the possible causes of the disappearance of the cowpat. They then use this knowledge to generate a hypothesis relevant to a realistic situation (Jim complains that there is always more cow dung on his concrete driveway than on the fields. He says the cows must prefer to drop their dung onto the drive. Can you think of any

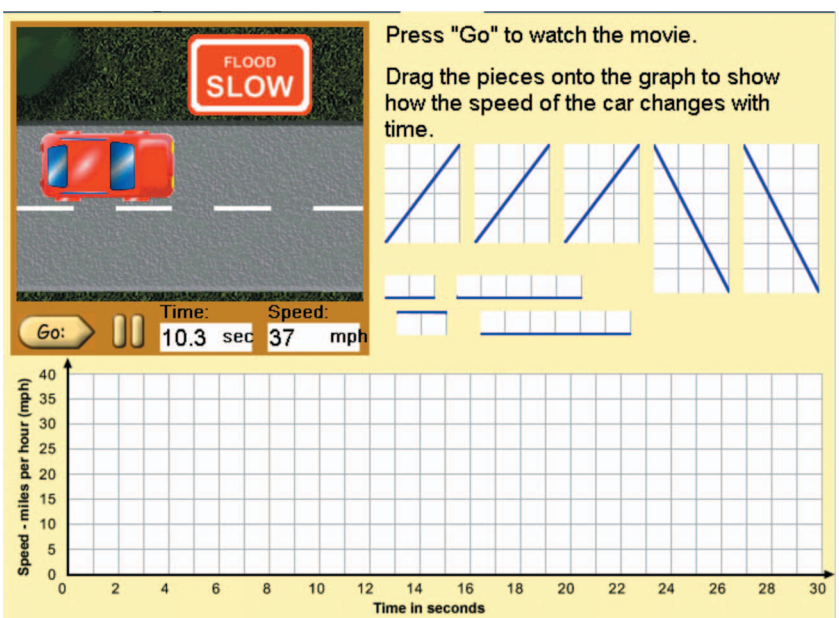

FIG. 3. Cowpats.

better explanation?). Here, the role of the computer is to provide a dynamic display involving multiple variables, which students use to generate and test hypotheses.

Discovering Rules

Computer microworlds can be established in which students have to discover the rules which control different aspects of the situation. These rules can be algebraic relationships, or functional rules ('this button doubles the number of lights that are switched on'). These tasks reward systematic working, and hypothesis testing. *Street Lights* (Figures 4a–4c) is a task for 13-year-olds and provides an example of a task where students have to discover some simple rules.

Finding Relations

Here, students are asked to discover empirical rules about how a dependent variable changes as a function of one or more other variables. Problems can be set in the context of pure mathematics, in a science laboratory, or in some applied setting.

Chasm (Figure 5) is a task for 13-year-old students which allows them to vary the width and thickness of a plank supported on two scaffolding poles, and to see the effects on the load the plank will bear. They are asked to write an equation which allows the load-bearing capacity of the plank to be calculated for any combination of width and thickness.

Cogs (Figures 6a–6c) is a problem designed for 13-year-olds where students have to discover the relationship between the number of cogs on meshed cog wheels, and the relative number of rotations. From easy starting examples where they can experiment with every combination of cog wheels, students are required to discover a general rule which can be used to solve problems in a new situation.

Testing Solutions

A critical part of problem solving is the process of testing possible solutions against the original task demands. Key components of this process include:

- seeking completeness and rigour;
- generalising rules;
- proving.

Completeness and rigour. Traditional methods of teaching encourage students to move on to the next problem, once they have produced a single successful solution to the current problem. In many real-world situations, this is inappropriate—exemplified by diagnosis and remediation in spheres such as medicine and industrial process control, where it is important to find all of the faults in a system. Some of our tasks require students to produce exhaustive solutions—to find all the mistakes in a design, to show why they are errors, and to suggest remedies. *Cogs* (Figures 6a to 6c), for example, requires two different solutions to the same problem.

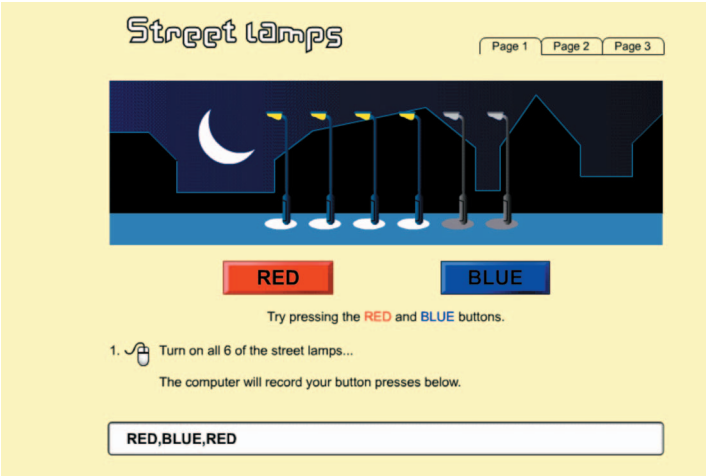

FIG. 4a. Street lights-page 1.

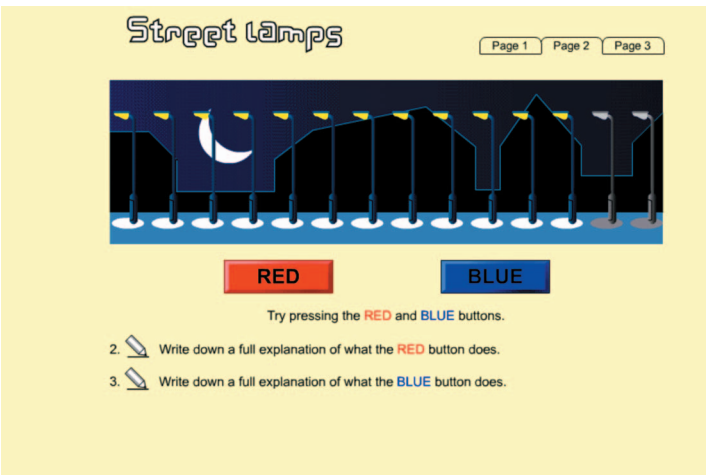

FIG. 4b. Street lights-page 2.

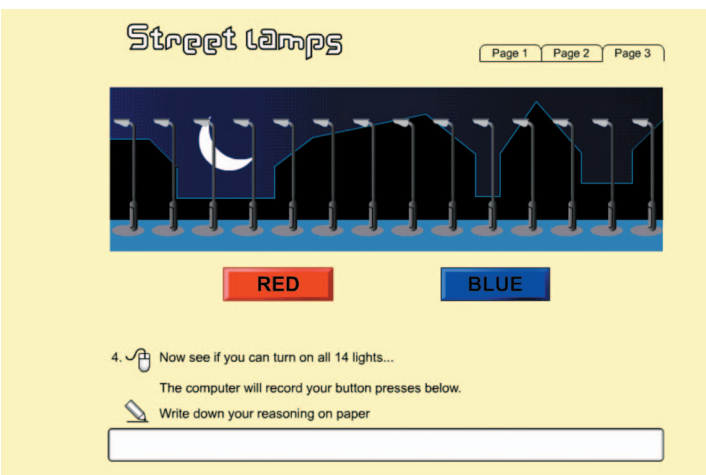

FIG. 4C. Street lights-page 3.

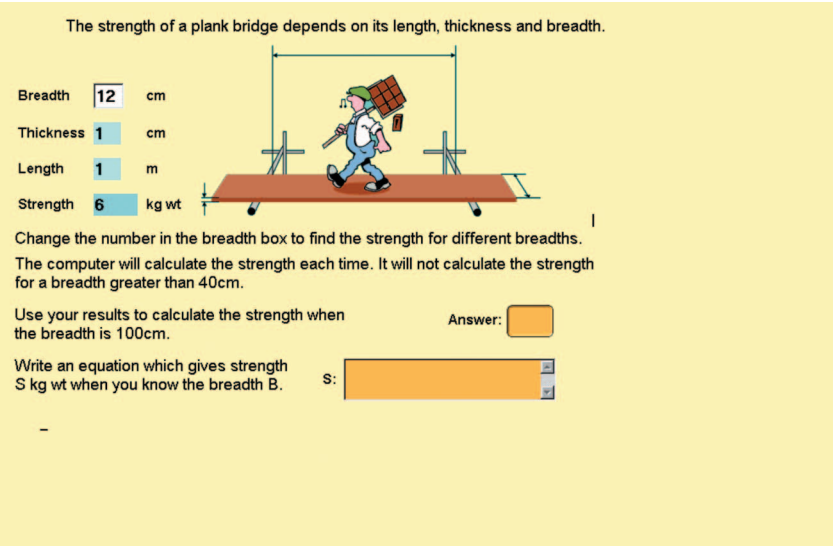

FIG. 5. Chasm.

Generalising rules: Pyramids (Figure 7) requires 13-year-old students to determine the relationship between one number and another. Early parts of the problem can be completed by trying specific cases. Later parts require students to create an algebraic relationship between the numbers, and to use this to solve a problem. The role of the computer is to provide a microworld which students can explore systematically, in order to discover a rule, which they then use to solve a problem outside the range of cases supported by the display.

Explaining Solutions

Problems can have a variety of solutions. In mathematics and science, there is a tendency to present problems which have just one solution. This conveys unfortunate messages about the nature of mathematics and science, especially in realistic contexts, where decisions must be made about problems which have many possible solutions. We set tasks which require 'constrained decision making' to assess relevant skills.

Holidays (Figures 8a–8d) is a task for 9- and 13-year-old students; it casts the student in the role of a travel agent offering advice to clients about possible holiday destinations, given their wishes and constraints, such as time of travel, costs, and the sort of holiday they want. Students use a database of holidays to narrow the choice down to just two or three possible venues. They are then asked to write a recommendation to the client about where to go. The marking scheme rewards answers which identify appropriate holidays, but in addition compares acceptable alternatives based on secondary (unspecified) criteria contained in the database, such as information on local sites of archaeological interest.

Here, the computer is providing a wealth of data which students must search

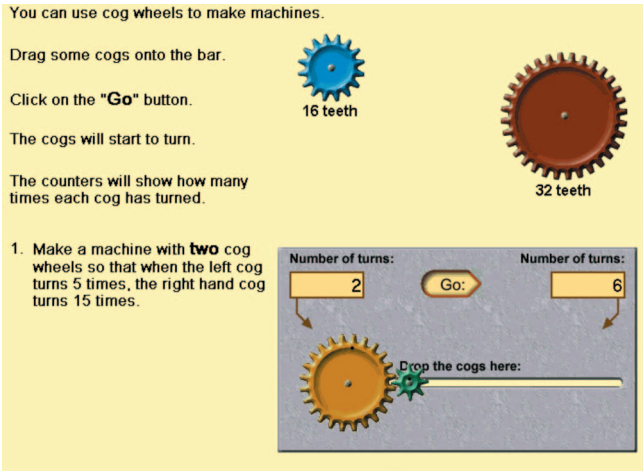

FIG. 6a. Cogs-page 1.

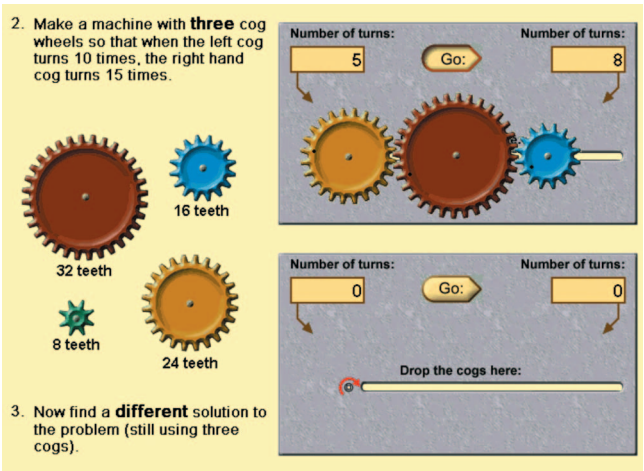

FIG. 6b. Cogs-page 2.

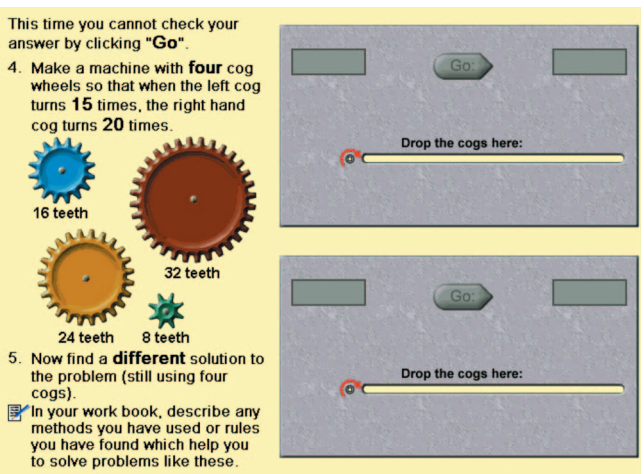

FIG. 6c. Cogs-page 3.

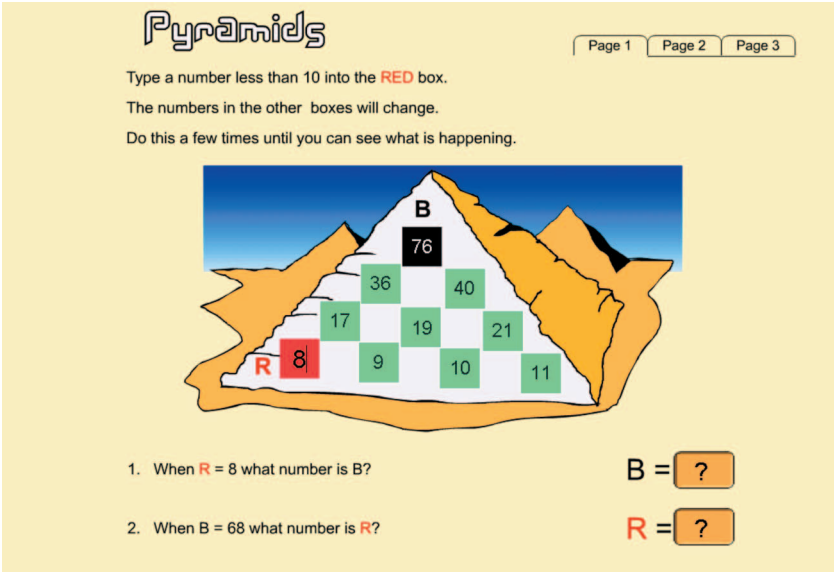

FIG. 7. Pyramids.

through to satisfy particular constraints. Students are given a realistic role as a problem solver; searching the database to find an acceptable solution actually finds at least two solutions. A good response from the student is to provide the 'client' with the evidence, and to suggest the circumstances in which one acceptable solution might be a better choice than another.

Science and mathematics are often used for decision making about key aspects of all our lives. In tackling any problem in a realistic setting, a variety of approaches can be taken. The evaluation of solutions often depends on moral and social values rather than the resolution of technical issues. Tasks must be sensitive to such issues.

Fishing (Figures 9a–9c) is a task for 13-year-old students which presents a problem where the student is asked to make recommendations about the number of boats that should be used by a company intent on making money by renting boats out for fishing. Data are available which predict the pattern of fish catches over a ten-year period, for different numbers of boats. Students are asked to use these data to make recommendations about the number of boats. If a small number of boats is used, fish stocks remain stable, but revenue is small. If a large number of boats is used, catches (and revenue) are high initially, then fall dramatically as fish stocks are depleted. The scoring scheme rewards both answers based on the idea of setting the number of boats at the highest level consistent with maintaining the fish stocks, and an answer which suggests the owners use as many boats as possible, then sell the business after one year, using the very high fish yields as part of the marketing.

Here, the role of the computer is to present sets of data in ways which students find very easy to use. The graphic interface allows the students to view large data sets very quickly. More interestingly, the dynamic display seems to allow students to

FIG. 8a. Holidays-page 1.

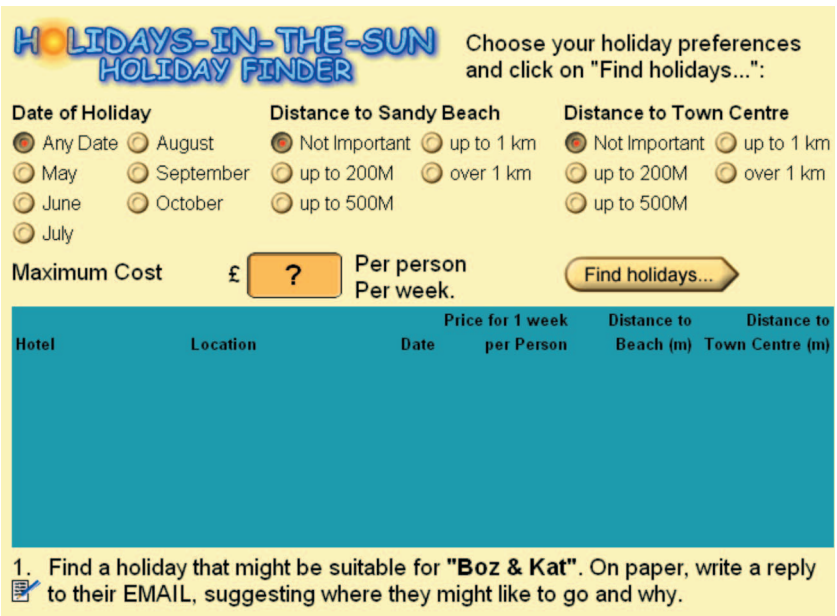

FIG. 8b. Holidays-page 2.—*continued overleaf*

represent the patterns in the data enactively. Some students will summarise the entire data set, mapping the decline in fish stocks under different conditions of depletion over a long time period, via a series of rapid hand and arm movements.

You are working for a holiday advice service and you receive the following EMAIL:

To:	advice@holidayhelp.co.wct
From:	Mills@rubynet.co.wct

We are a retired couple, trying to plan a week's holiday this summer. We would prefer a sunny, quiet resort with a sandy beach nearby. We are also interested in history.

We can afford up to £700 for the two of us.

We would be most grateful for your advice.

Yours,

Mr & Mrs Mills

Now go to page 4 and use the **Holidays In The Sun** website to search for a suitable holiday for Mr and Mrs Mills.

FIG. 8c. Holidays-page 3.

HOLIDAYS-IN-THE-SUN HOLIDAY FINDER

Choose your holiday preferences and click on "Find holidays...":

Date of Holiday: Any Date, August, May, September, June, October, July

Distance to Sandy Beach: Not Important, up to 1 km, up to 200M, over 1 km, up to 500M

Distance to Town Centre: Not Important, up to 1 km, up to 200M, over 1 km, up to 500M

Maximum Cost £ 350 Per person Per week. Find holidays...

Found 23 holidays:

Hotel	Location	Date	Price for 1 week per Person	Distance to Beach (m)	Distance to Town Centre (m)
Art	Es Cana	26-Sep	339	350	350
Art	Es Cana	05-Sep	339	350	350
Budak	Altinkum	21-Jun	319	500	500
Budak	Altinkum	12-Sep	299	500	500
Budak	Altinkum	05-Sep	299	500	500

Well maintained and traditionally furnished hotel in quiet residential area of resort. Altinkum has a wide stretch of golden sand and is only 4 km from the site of the Temple of Apollo at Didyma.

2. Find a holiday that might be suitable for **Mr & Mrs Mills**. On paper, write a reply to their EMAIL, suggesting where they might like to go and why.

FIG. 8d. Holidays-page 4.

Current Evidence

The tasks shown in the previous section were developed as part of The World Class Arena (www.worldclassarena.org). To date, four sets of live tests have been administered in the UK and elsewhere, each of which was preceded by extensive pre-testing. The evidence from these tests and trials is not yet in the public domain; here,

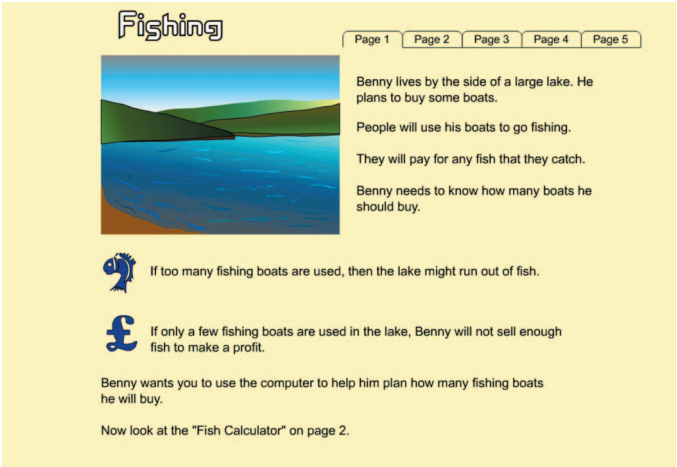

FIG. 9a. Fishing-page 1.

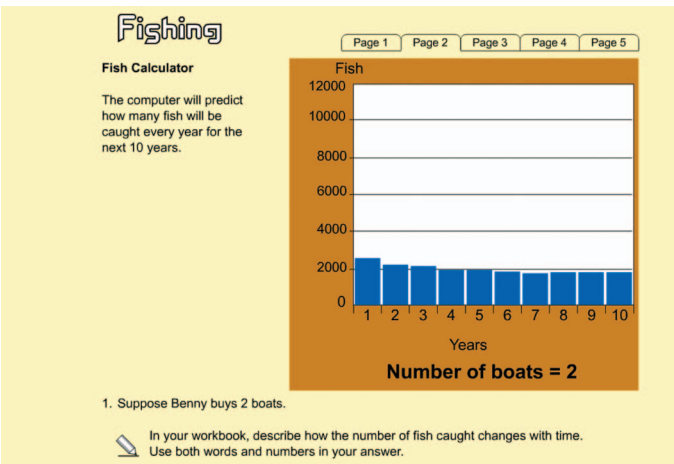

FIG. 9b. Fishing-page 2.

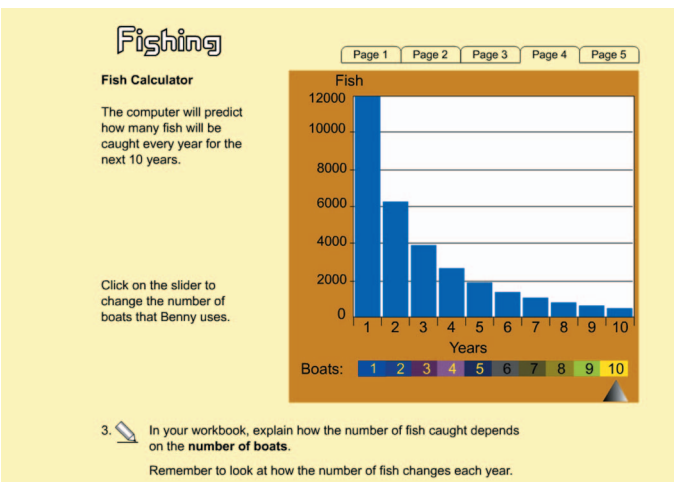

FIG. 9c. Fishing-page 3.

we offer an overview of the analyses conducted so far. Perhaps the most notable result is the ease with which students have interacted with computers. We (naturally) attribute this to the effort we put into interface design and testing. A more mundane explanation is that students have received a good deal of exposure to computers, and have experience of learning to use a wide variety of interfaces. The affective response from students is very strong—they really enjoy working on these tasks. This might be related to the sustained challenge the tasks present, which is similar to the reported reasons they like computer-based games (McFarlane *et al.*, 2002). The rest of this section considers cognitive aspects of the tests.

Students perform better on some tasks than we might expect—notably tasks that require them to reason from complex data sets (e.g. data with two independent variables and one dependent variable at age 9 years). We take this as a very positive sign that computers can play a leading role in the development of the skills which constitute the new educational agenda.

Data from live tests have shown that students in some schools perform dramatically better than students in some other schools. This suggests strongly that the skills we are trying to assess can be taught effectively, but receive little attention in a number of schools. Ongoing work involves observation in school, in order to understand the curricular practices which underpin the development of expertise.

Sex differences in performance are small, for the majority of tasks we have tried. Clearly, balancing tests so that they do not show sex bias is relatively easy if there are few sex differences on individual tasks.

The spontaneous use of good representations—such as tables, graphs or diagrams, or student creation of their own representations are rare.

Too often, work is characterised by guessing, rather than by a systematic attack on the problem. Unprompted use of systematic methods and systematic recording of results is rare.

Testing hypotheses is done well; creating hypotheses is done less well.

Students find it difficult to make generalisations (but this is done better at age 13 than at age 9).

Students find it difficult to create mathematical or scientific models of situations.

Completeness and rigour are rare. It is common for students to offer a single solution to a problem—or to identify just one fault when fault finding. This may well be a response to the testing situation, where students are used to moving onto the next problem once an (exactly one) answer has been found to the current problem.

On many tasks, students are able to show evidence of good reasoning skills. However, explanations are often weak. It is not clear whether this reflects a cognitive problem, or reflects students' beliefs about the 'didactic contract'. In mathematics and science, reasoning and the creation of a correct answer can be seen to be all that is required—the need for students to produce clear explanations is not seen as an important part of performance.

Conclusions

We have made considerable progress in the development of tasks and tests which

reflect important aspects of the new educational agenda. A number of technical problems concerning interface design, and the delivery and return of test data have been solved. Students respond very positively to the tests we have developed.

Computers make it easy to present new sorts of tasks, for example tasks where dynamic displays show changes in several variables over time, or which present video of a situation which students must model. A wide variety of representations can be supported, and students can be asked to switch between them. The interactive properties of computers make them well suited to the assessment of process skills—problem exploration, discovering rules, finding relationships, developing effective strategies—by the use of simulations, microworlds and interactive games.

Using computers to give students control over how data is presented allows them to work with complex data sets of a sort that would be very difficult to work with on paper. This allows students to be set tasks in realistic contexts, using realistic data to address real problems of considerable complexity, using resources and methods that are familiar to professionals working in the relevant field.

Student performance shows some surprising successes in handling complex data, which supports an optimistic view of the potential of IT to promote intellectual development. It also reveals some lacunae in a number of aspects of performance, such as spontaneous use of different problem representations, systematic working and recording, hypothesis generation, and clear explanation.

Large performance differences between students in different schools suggest that many of these aspects of performance can be taught successfully.

A Technical and Conceptual Addendum

The WCA tests are designed to assess problem solving, and problem solving can be characterised as 'knowing what to do when you don't know what to do'. This means that tasks must be non-routine in nature. If WCA tests are to have a positive impact on educational practice, they must provide a set of novel challenges on each testing occasion, for the obvious reason that if it becomes apparent that each test presents a set of familiar problems, then teaching students how to solve such problems will come to constitute a narrow curriculum for 'problem solving'. In contrast, if a test presents novel challenges, then the most effective method for test preparation will be to present students with novel problems, and to explore ways in which they might be solved.

The technical issue that arises from this analysis is the problem of test reliability. The sampling of tasks which present novel intellectual challenges, and which can cover topics in mathematics, all the sciences, and technology, might make the equivalence of tests taken on different occasions (which should be parallel forms in some sense) impossible to attain. Conversely, one might hope that, since every task is assessing some aspect of higher order thinking, we would expect performances on different collections of tasks to be highly intercorrelated.

The empirical evidence shows that student performances on different tasks, indeed, do correlate highly. At age 9 years, the average of the item-whole correlations over each of the five tasks in a single test was 0.59; and was 0.65 at age 13

years. Students take two tests as part of their assessment, and so (if we make the assumption that tasks on different tests are drawn from the same pool of tasks) we can be confident that if the same sample of students took WCA tests given on different occasions, their scores would be highly correlated.

REFERENCES

ADEY, P. & SHAYER, M. (1994) *Really Raising Standards: cognitive intervention and academic achievement* (London, Routledge).

ANDERSON, J. R. (1993) *Rules of the Mind* (Hillsdale, NJ, Lawrence Erlbaum).

ARCHENHOLD, W. F., BELL, J., DONNELLY, J., JOHNSON, S. & WELFORD, G. (1988) *Science at Age 15: a review of APU findings 1980–1984* (London, HMSO).

BADDELEY, A. D. (1990) *Human Memory: theory and practice* (Hove, Lawrence Erlbaum).

CLARKE, D. & PETER, A. (1993) Modelling teacher change, in: ATWEH, B., KANES, C., CARSS, M. & BOOKER, G. (Eds) *Contexts in Mathematics Education*. Proceedings of the 16th Annual Conference of the Mathematics Education Research Group of Australasia, Dates?, Brisbane. Q1

DAVYDOV, V. V. (1995) The influence of L. S. Vygotsky on education theory, research, and practice, *Educational Researcher*, 24 (3), pp. 12–21.

DOWNES, T. & ZAMMIT, K. (2000) New literacies for connected learning in global classrooms, in: H. TAYLOR & P. HGENBIRK (Eds) *Information and Communication Technologies: the school of the future* (London, Kluwer Academic Publishers).

FLAVELL, J. H. (1976) Metacognitive aspects of problem solving, in: L. RESNICK (Ed.) *The Nature of Intelligence* (Hillsdale, NJ, Lawrence Erlbaum).

FULLAN, M. (1991) *The New Meaning of Educational Change* (New York, Teachers College Press).

MCFARLANE, A., SPARROWHAWK, A. & HEALD, Y. (2002) *Report on the Educational Use of Games* (London, Department for Education and Skills). Available at www.teem.org.uk

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS (NCTM) (2000) *Principles and Standards for School Mathematics* (Reston, VA, NCTM).

RIDGWAY, J. (1997a) Vygotsky, informatics capability, and professional development, in: D. PASSEY & B. SAMWAYS (Eds) *Information Technology: supporting change through teacher education* (Chapman Hall, London).

RIDGWAY, J. (1997b) Why measure development? in: L. SMITH, J. DOCKRELL & P. TOMLINSON (Eds) *Piaget, Vygotsky, and Beyond* (London, Routledge).

SCHOENFELD, A. H. (1985) *Mathematical Problem Solving* (Orlando, FL, Academic Press).

SCHUBERT, S. (2000) The impact of modelling in education of informatics on collaborative learning with school Intranets, in: H. TAYLOR & P. HOGENBIRK (Eds) *Information and Communication Technologies: the school of the future* (London, Kluwer Academic Publishers).

TAYLOR, H. G. (1996) The TTACOS Project—Laying the Foundation for National Technology Standards for Students in the United States, in: D. PASSEY. & B. SAMWAYS (Eds) *Information Technology: supporting change through teacher education* (London, Chapman Hall).

VYGOTSKY, L. S. (1981) The genesis of higher mental functions, in: J. WERTSCH (Ed.) *The Concept of Activity in Soviet Psychology* (New York, Sharpe).

WEBB, N. L. (1999) *Alignment of Science and Mathematics Standards and Assessment in Four States* (NISE Research Monograph No. 18) (Madison, WI, University of Wisconsin-Madison).

WOLF, A. (2002) *Does Education Matter? Myths about Education and Economic Growth* (Harmondsworth, Penguin Books).

WOLF, A. & TIKLY, C. (2000) *The Maths We Need Now: demands, deficits and remedies* (London, Institute of Education).

Author Query Sheet

Manuscript Information	
Journal Acronym	cAIE
Volume and issue	10(3)
Author name	Ridgway
Manuscript No. (if applicable)	5

AUTHOR: The following queries have arisen during the editing of your manuscript. Please answer the queries by marking necessary corrections **at the appropriate positions on the PROOFS. Do not answer the queries on the query sheet itself. Please also return a copy of the query sheet with your corrected proofs.**

QUERY NO.	QUERY DETAILS
1	Please give date of conference