A New Model of the Question Answering Process

Alastair Pollitt & Ayesha Ahmed

Research & Evaluation Division

University of Cambridge Local Examinations Syndicate

A paper presented to the International Association for Educational Assessment, in Bled, Slovenia, May 1999.

Introduction

The central phenomenon of the whole examination process is what happens when a candidate meets a question. This is the focus of all our activity: no amount of good administration, good teaching or wise judgement can compensate if there is something wrong with the questions. Educational examining is a form of mental measurement or of measuring some property of each candidate's mind, using as evidence the product they generate when they meet each question. Understanding assessment therefore depends fundamentally on understanding what happens during this process - as a candidate tries to answer a question.

Validity

It seems generally accepted today that the essential form of validity is that called *construct validity* (Messick, 1988). The view taken here is that the construct in question is in reality a *psychological process*, the combination of mental activities that are required to perform at a certain level of proficiency on the test in question. Thus, for example, the construct we measure in a language test is in reality the set of language-related, cognitive, and other processes that students use in answering the questions that make up the test: the validity of the language test then depends on how closely these processes correspond to those involved in the sort of activity we understand as language use. Unlike some views of validity, this is essentially *intrinsic*; given an understanding of the activity the test is meant to assess, validity consists in managing the assessment procedures in such a way that candidates' mental activities during the test will correspond as closely as possible to the mental activities of a person engaged in real life use of the knowledge or subject being assessed. (Related conceptualisations of validity can be seen in Bachman's (1989) notion of *interactional authenticity*, and Macnamara's (1996) notion of *strong performance testing*.)

The motive for the research reported here is the quest for a better understanding of the question answering process, and so of the construct being assessed. Here *question* refers in general to the range of written questions we commonly set in educational examinations. It may be worth stating at the outset one exception to this definition - we have found it extremely difficult to model the process of answering multiple choice questions, and are inclined to think that, perhaps for this reason alone, they are of questionable validity for educational assessment . *Answering* refers to the whole activity of a student with respect to a particular question including, to some extent at least, their prior experience of learning the relevant subject, while *process* indicates that our interest is not primarily in the product (answer) produced but in understanding how it came about, even though most of our evidence may reside in the answers. Our story of modelling the process begins 20 years ago.

What makes exam questions difficult?

In the traditional approach to construct validation variables were sought that correlate with students' test scores; students scoring high on the test should also score predictably high on certain other, supposedly related measures, but unpredictably on less related ones. Stenner et al (1983) argued, in contrast, that knowing why one question is harder than another is essential to knowing what it is that your test is measuring, and that evidence about the relative difficulty of questions is more reliable (because more data is available) than evidence about students. Simultaneously, our research team (in Edinburgh University) adopted similar principles in a study of question difficulty that aimed to improve the construct validity of school examinations (reported in Pollitt et al, 1985).

The project

Five subjects were chosen: geography, mathematics, chemistry, English and French (ie a first and a foreign language), and question papers from the Scottish O Grade examinations of 1980 were studied, together with responses to them from students of varying ability. Sources of difficulty and easiness were identified and a classification scheme developed to help question writers and scrutineers understand, and avoid, the problems that had been identified. Of particular relevance to the present paper is the model that was developed to try to explain the difficulty of reading comprehension questions in the two languages.

The language comprehension model

Reading comprehension in these examination papers (sometimes called *interpretation*) was assessed in the common format of a text followed by questions to be answered in a phrase or sentence. At its simplest the model is a sequence of five stages whose order is more or less logically necessary:

Students begin by reading, and 'comprehending', the given text. Then, they

1 *understand* the question,

2 *search* the text for the relevant part (or parts),

3 *interpret* the parts of the text, and

4 *compose* the answer.

So far, logic demands the model since, for example, it is not possible (or should not be possible!) to find relevant text without having understood the question. Step 3 was subdivided to separate the decoding of the text from the process of generating a response to the question based on that understanding, a distinction which roughly corresponds to the familiar linguistic separation of 'literal' meaning from 'inferred' meaning, though it is not necessary that these two parts of the process will happen in a given order. The five parts were then expanded into a sequence of 44 steps, based on empirical evidence from script errors and interviews with students. A partial confirmation of the model was obtained from a sequential multiple regression analysis of question difficulty using subjective ratings of the step difficulties (Pollitt & Hutchinson, 1987).

Analogical reasoning

About the same time, Sternberg (1977) was engaged in studying inferential reasoning processes in more formal settings. He proposed a model of the psychological processes involved in solving analogy problems, of the form *A is to B as C is to ...?*, which involves six "components":

encoding the left half of the analogy is 'read' into working memory, (the words A and B are 'read'),

inference the relationship is inferred (the meaning of the problem is constructed),

mapping	the relationship is mapped on to the right half of the analogy (a similar structure to that of the problem is sought in long term memory),
application	the mapping is applied, forming an image of the ideal solution,
	(just for a multiple choice context an extra component called *comparison* is inserted here)
justification	the actual solution is selected and justified,
response	the response is generated.

Although described as "components" it is clear that these are intended to form a sequence which, in simple cases at least, will represent how the problem is actually solved. Like Pollitt et al, Sternberg recognised that particular instances of analogical problem solving will often involve repetitions and cycles of the components.

The language comprehension model (2)

An important development in the model of reading comprehension came with the realisation that the presence of the text during a reading test may invalidate the test (Taylor, 1994; Pollitt & Taylor, 1996/7). In 'real' reading the text is processed, and the product of this processing is an internal mental representation of the text. In a valid reading test students should answer questions using this representation rather than the original text. The model becomes:

Students begin by reading, and 'comprehending', the given text. Then, they

1. *understand* the question,
2. *search* the mental representation for the relevant part (or parts),
3. *interpret* the parts of the mental representation, and
4. *compose* the answer.

With this change the model of language comprehension is easily generalised to educational learning in general, as will be explained below.

The question difficulty project

The new model to be reported here is based on work begun in 1996 in Cambridge. Most of the work has been funded by UCLES, with some additional support from The Qualifications and Curriculum Authority of England. Like the Edinburgh project, five subjects were chosen: GCSE geography, mathematics, combined science and French, and O Level English (a foreign and a second language, because first language examinations in the UK today place more emphasis on literature than in 1980). Most reports written so far from this work have been internal and confidential, but reports on each of the subject studies are being prepared for publication. These describe the Sources of Difficulty and Easiness (SODs and SOEs) identified in each subject from analysis of scripts, interviews and protocol studies carried out with students, and the results of experimental studies in which modified questions were compared with the originals in confirmatory tests of the proposed SODs and SOEs. Some of these results have been reported to IAEA in Hughes (1997) and Ahmed & Pollitt (1999).

Cognitive psychology has changed greatly in the last 25 years, and a new model of the question answering process is needed to reflect current understanding. We begin with a list of particular fields of research within psychology that bear most on the answering process.

Relevant domains of cognitive psychology:

Learning

Since the avowed purpose of examinations is to assess a student's levels of knowledge and skill, it is obvious that an understanding of the learning process will be of interest to research into answering processes. We may formulate the question to be addressed (using *knowledge* as a generic term to include *skill*):

How is knowledge acquired?

Memory

This question leads into the very active field of memory research. Students certainly recognise the importance of memory in examinations, but to a psychologist the questions of interest are:

How is knowledge represented in the brain?
How is knowledge accessed?

Problem solving

This is a traditional area of interest for educational psychologists, asking the question:

What strategies are used to solve problems?

We believe that this work has over-emphasised deliberative thinking, and has tended to miss an important aspect of the process, which might be summed up in the question:

How do students identify relevant knowledge?

Attention

A student can only deliberate about concepts that are active in working memory. This raises questions about how much content can be considered at a given time. A particular feature of assessment that has not been widely researched, at least from the point of view of cognitive psychology, is the effect of cognitive stress on examination performance. Both the perceived importance of the outcome (the so-called high stakes) and the time limits imposed are likely to make students behave differently in examinations from how they would behave in other circumstances. What effects will these have on students' ability to attend to relevant features of problems? While a certain amount of arousal may optimise performance, the main question is:

What constrains students' thinking under examination conditions?

Language comprehension

Some of the most significant developments in recent psychology concern our understanding of the cognitive processes involved in understanding language. With this, we have come to realise how central language comprehension is in examining - after all, nearly every task is communicated to the candidate through the medium of language. This question is not trivial:

How do students understand the question/problem?

Furthermore, the models now used to describe understanding of language are special cases of models for thinking and learning, and an understanding of language comprehension can be generalised to those domains.

Language production

It is then, of course, necessary to consider the other end of communication, for in the vast majority of cases students communicate their responses to the examiners through the same medium of language. There are well known difficulties in the writing process, that persist long after the reading process has been more or less mastered, and we must ask:

How do students compose a (written) answer?

The new model

The key to extending the Edinburgh model from reading comprehension to all examination questions is to see that the *text* can be replaced by the *subject*. Understanding a subject is just a very large version of understanding a story, with many different kinds of input - seeing, reading, hearing, etc - in place of just reading the text. Then the process of *searching* the text or the representation provoked by the text, as described earlier, simply translates into *searching* through one's understanding of the subject for relevant knowledge. The sub-step of *interpreting* the piece or pieces found translates into specific understanding of that subtopic in the subject, and *constructing* the answer becomes the more general matter of generating a solution to the problem. If we include learning the subject as Step 0, the model then is:

0	*Learning*	learning the subject
1	*Reading*	understanding the question
2	*Searching*	accessing relevant aspects of memory
3	*Interpreting*	re-interpreting stored knowledge to match the question
4	*Solving*	generating a response to the question
5	*Composing*	writing an answer for the examiner to read

It is still proposed that these processes operate more or less in the sequence given, although it is recognised that repetitions and cycles will occur. It is particularly important to understand the difference between Steps 0 and 3. For any question, a candidate's response will depend on the understanding they have *at the time of answering*; since remembering always involves reconstructing, the immediate understanding can be considered as being a reconstruction just to respond to the question. It is certainly possible to make errors in recalling a topic that is 'well understood'.

This model can be seen to parallel Sternberg's analogical reasoning model quite closely:

	Cambridge	**Sternberg**
0	*Learning*	
1	*Reading*	*encoding* and *inferring*
2	*Searching*	*mapping*
3	*Interpreting*	*application*
4	*Solving*	*justification*
5	*Composing*	*response*

Current theory would see the construction of the problem's meaning as simultaneous with input, rather than following it, and so *encoding* and *inferring* are both part of what we label as Reading. Sternberg's notion of *mapping* clearly involves searching 'learning' or 'memory' for similarities, and his discussion of *justification* seems to correspond quite closely to what we mean as Solving.

We may also locate each of the psychological domains listed earlier as being primarily concerned in one of these phases:

	Phase	**Domain**
0	*Learning*	Learning
1	*Reading*	Language comprehension
2	*Searching*	Memory
3	*Interpreting*	Attention
4	*Solving*	Problem solving
5	*Composing*	Language production

The value of a model like this is that it provides a framework within which to consider the relevance of psychological research. An examiner can then systematically consider how students are likely to react and respond to any given task that is set in an examination with - we hope - consequent improvements in examination quality.

The components

This section will outline relevant recent research in the six domains. Our UCLES' research is mainly concerned with exploring the relevance of these theories to the matter of answering examination questions, but the review here covers the theories themselves. The presentation of this paper will be illustrated with examples to show the relevance to our concerns. The review is, of course, very brief and incomplete, and we are continuing to explore all of the areas mentioned.

The sub-sections below use the phase numbers from the table above. It is helpful, however, to begin with language comprehension since, as mentioned earlier, it provides a simpler basis for understanding the whole of learning.

1 Language comprehension

How do students understand the question or problem?

A feature common to almost all theories of reading (and listening) comprehension in recent decades is to see the elements of understanding as **personal** and familiar, selected by readers from their unique repertoire. Most theories employ a metaphor of **construction**, which portrays the reader as building meaning from 'prefabricated units'. Two slightly different camps may be identified, first the *AI* group and then the *psychologist* group:

Minsky (1975) - a *frame* is "a remembered framework to be adapted to fit reality, by changing details as necessary"

Schank & Abelson (1977) - described *scripts*: sequential frames for routine activities

Sanford & Garrod (1981) - scenarios: situation-specific frames

Bartlett (1932) - first described a reader's *schemas* in reading comprehension

Anderson (1977) - schemas are "ideational scaffolding"

van Dijk (1981) - schemas are "higher-level complex (and even conventional or habitual) knowledge structures"

Johnson-Laird (1981) - words are "cues to build a familiar mental model"

The psychologists see the reader as an **active** participant, in contrast to the passive role implied by the AI group, and even more in contrast to the 'receiver' or 'decoder' of earlier information transfer models of communication. In addition, they emphasise that the linguistic input is a **prompt**, provoking the reader to do something, rather than controlling what the reader does.

Mental Models

The stance we adopt is based on the mental models approach of Johnson-Laird, elaborated for language processing by Garnham and others. It is common nowadays to use the phrase **mental representation** rather than mental model to avoid seeming to imply special importance for the visual processing mode. Equally, the linguistic mode is not preferred: "mental representations usually model aspects of the world rather than aspects of linguistic structure" (Garnham, 1987). The last two quotations illustrate well the essence of the approach. The reader actively tries to construct a coherent mental representation in their mind of some real or imaginary world. This representation is almost entirely built out of elements - ideas, concepts, meanings - that they already possess, which are brought to their attention as a result of stimulation by the text they see. The resulting representation *is* their understanding of the text.

A mental representation will include relevant **schemas**, which are particularly large and stable elements in the reader's mind. A schema is a pre-fabricated framework which typifies a certain context or kind of event; we have schemas for social conversation, for buying groceries, and for sitting examinations, as well as smaller ones for various kinds of reasoning. **Scripts** can also be seen as a certain kind of schema, containing a set of participants and actions, and an order in which they are likely to occur. One important feature of schemas for our purposes is that they are frameworks of expectations, and candidates' behaviour in an examination setting will be strongly affected by their expectations of what they ought to be doing.

Context is a particularly powerful influence on performance, largely because it activates schemas so readily. If we remember that the mind we are examining spends most of its time living in the real world, and some of its time even studying other subjects, we may better realise how influential contextual effects and expectations can be.

Of particular importance is the *ordo naturalis*: events often have a natural **order** and we expect a text to reflect it. We normally assume that the order of events in the text will reflect the real order of the events. We also assume that information we are given is relevant (Grice, 1967) and so are quick to assume causality as well, even though the text may not state it and the author not intend it. We also know that flashbacks and other loops make text comprehension more difficult.

For narrative the natural order is chronological: it is less clear what is 'natural' for the various kinds of non-narrative text, but unexpected order in a text or question will still be a source of difficulty in examinations.

Language processing is a **real time** process. Even though a reader can re-read a piece of text it seems clear that the mental representation will mostly be determined by the first reading. A text is read in a particular sequence and the ideas provoked by the words are built on to the current mental structure: if and only if the reader cannot make new input fit coherently, then a new structure is started (Gernsbacher, 1990). A text can influence this process by careful, or careless, paragraphing and by using the various discourse markers that language offers.

A further very important issue for us concerns **salience**. Most reading is not carried out with great care; reading provokes expectations for what the text is going to contain next, as well as what the 'world' we are modelling will contain. We read selectively, using our intelligence to decide what to pay most attention to. Words of particular salience - perhaps because they involve emotions, or are especially striking, or unexpected - will force themselves to our attention and may come to dominate the model we construct.

0 Learning

How is knowledge acquired?

In evolutionary terms it would not have made sense to develop a special system for comprehending language input when a perfectly decent system already existed for making sense of auditory, visual and any other sensory input. Therefore it is no surprise that Learning is 'like' language comprehension, but on a larger scale. The input is multi-modal, rather than just linguistic, and consequently much richer.

This has the important consequence that each pupil's learning is **idiosyncratic**. They construct their own understanding of, say, geography from their own particular experiences. They each read and watch different books, magazines, television programmes, and see different geographical phenomena in the real world; they find different bits of the subject particularly interesting or boring, for reasons we cannot fathom; they remember particular incidents during their learning which for some unpredictable reason were particularly salient. Their mental representation of geography will contain their own unique mix of memories, understandings, misunderstandings and rationalisations, with sights, sounds, smells and emotions all strongly influencing the significance and the accessibility of each part.

We test the students against an official understanding of the subject, and most of the time we do it through the medium of writing. Therefore, crucially for students, examinations involve the translation of aspects of mental representations from idiosyncratic multimodal structures to the **authorised,** linear, textual, version.

2 Memory

How is knowledge represented in the brain?

Psychologists distinguish various types of memory, such as semantic and episodic, or declarative and procedural, but our concern in examinations is almost wholly with memory for meaning, ie **semantic** memory. Concepts constitute the most important basis of semantic memory, of thinking, and of language. How are concepts stored? Many are stored hierarchically.

A concept hierarchy

commerce
- business
  - SHOP
    - grocery
      - delicatessen

A concept like *shop* belongs in a hierarchy, since it is more specific than *business* and *commerce*, but more general than *grocery* and *delicatessen*. Apparently in accordance with a principle of economy, each concept 'inherits' all the properties of higher concepts and passes them on to subordinate concepts - unless they are specifically over-ridden.

Many concepts depend strongly on **language**, to the extent even that we may know them best as verbal definitions. This is particularly true of technical concepts, and is perhaps most problematic when a word is used in both a technical and an everyday sense. Students must then avoid the default everyday sense and avoid whatever connotations the everyday sense may carry.

There has long been debate about whether a concept is defined by a set of features or as an essence. The current view for most concepts is that they are best seen as fuzzy sets, with a clear core and vague limits. Thus a **prototypical** cat will be small, domestic, and friendly, with four legs, whiskers and so on. But losing one leg will not disqualify it from being a cat; nor will its being large, wild or fierce.

How is knowledge accessed?

Perhaps the most influential theoretical treatment that deals with memory access is the ACT* theory of Anderson (eg 1983). It has three essential features: **memory**, **association** and **activation**.

Mental activity consists of 'programs' operating on memory, often causing changes to it. Any thought creates or strengthens a node in memory, associating two concepts temporarily or permanently. This is illustrated by the diagram in the handout showing how a reader would process *Sarah ate breakfast*.

How do students identify relevant knowledge?

This is mostly an unconscious process, depending on **activation** and **salience**. Activation means that a particular concept is made available to the student's working memory or, in informal language, is brought to mind. Incoming words are "cues" that provoke activation of certain concepts and words, and these in turn raise the activation level of all the concepts and words they are associated with, irrespective of the relevance of the association to the problem under consideration. More salient words are more likely to result in significant activation of their associated concepts than less salient ones.

Searching and Matching

Searching is an automated process of seeking patterns of activation that correspond, more or less well, to the problem as the student understands it. Matching is also automatic, but its results will be monitored consciously for relevance as part of the Solving process.

Stereotypes

The relevance of these ideas to examinations is largely to do with the automatic nature of the activation and searching processes. For reasons of efficiency thinking is mostly an automatic process guided and monitored consciously - with more or less care - by attention. In examinations attention may fail to monitor effectively for several reasons.

Consider the conceptual hierarchy again. In most such structures one level can be agreed as the **base level** (*shop* in this case). The base level may be defined informally as the most familiar and distinctive level in the hierarchy, and so the one with which we reason with most ease. If we are asked to reason at other levels this will necessarily involve inference upwards and downwards from the base level.

Base level thinking involves **stereotypes**. The stereotypic 'shop' will probably be rather a personal one, quite different in Singapore or the United States from in Europe. It functions as a fairly concrete referent for what is actually a fairly abstract concept. Stereotypic thinking is more efficient - most of the time. The dangers, though, are obvious. At any level in the hierarchy there is a tendency to use one stereotypic example to represent the whole concept. Thinking consists of making connections between concepts, and all of the properties of the stereotype will be activated and may be applied inappropriately to a particular example.

3 Attention and stress

High stakes examinations are notoriously stressful, in the everyday sense of the word, but here we mean that candidates have to deal not only with the cognitive demands of the questions set but also with other demands on their attention. Time is usually limited, familiar reference materials and other aids are usually absent, normal activities like consulting others are usually forbidden, and natural behaviour is restricted in various ways; all these combine with affective phenomena such as anxiety and fear of failure to distract the student and most probably to reduce the quality of the performance.

What constrains students' thinking under exam conditions?

Several areas of study may contribute to understanding this issue.

Pre-attention

Wason & Evans (1975) proposed a dual process theory of reasoning, and Evans developed this to explain patterns of error in various cognitive tasks (Evans, 1989). While one of these processes is the traditional set of deductive and other logical inferences, he points out that these involve **attention** to particular features of the task. Logically, then, there must be a **pre-attentional** process by which these feature are unconsciously selected for attention. This phase (part of *Reading* in our model) is sensitive to biasing influences of the text, the context, or the student's own prior knowledge. It seems likely that examination conditions will reduce students' ability to avoid bias of this kind.

Channel capacity

Only a certain number (usually said to be about seven) concepts or relationships can be retained in working memory at one time. Others will be activated, and so available for recall. Under stress, this capacity is reduced, as some attention is distracted to coping with the stress.

Proceduralisation

One important way that experts differ from novices is in their ability to proceduralise their skills, so that what is for novices a complex set of operations is for the expert a single procedure. This, of course, increases the amount of knowledge simultaneously available to attention.

Closure and checking

Experts under stress seem able to retain an ability to monitor their thinking; novices seem to stop too quickly, as soon as a conclusion is reached, without ensuring that the conclusion makes sense.

Suppressing stereotypic properties

Similarly, novices seem to reason more with stereotypes, and under stress may be less able to keep monitoring for exceptional, non-stereotypic, characteristics of the particular case in the question.

4 Solving

This is a well researched and reported area in cognitive and educational psychology. We deal here only with a few general features of recent approaches.

Planning - monitoring - evaluating

Most descriptions of 'strategy' for problem solving involve three aspects called something like:

planning - identifying goals, breaking them down into sub-goals, identifying promising routes;

monitoring - ensuring that current activity is consistent with the plan;

evaluating - 'measuring' the 'distance' from the current goal.

Such a model applies best to large scale tasks which require deliberate planning. But most examination questions seem to involve processes that are much more automatic than this: how does the mind work when we are less conscious of our strategies?

What strategies are used to solve problems?

Sternberg's (1982) description of the components of problem solving in the case of analogies shows that the whole process may be analysed recursively. By mapping his model to ours we are identifying step 5 - called justification - with the core of problem solving. This may at first sight seem strange; surely justification is a *post hoc* activity, used to check the solution that 'problem solving' has given us? Consider, however, Johnson-Laird's (1980) insistence that, in the mental model approach, there are no rules of inference. Reading, and thinking, and remembering, all involve constructing models of a 'world', and it is this construction that automatically 'creates' the answer to the problem. The student's task is to recognise potential solutions (through searching for a match) and then to confirm it (through justification).

The concept of **problem space** is useful for discussing the effect of manipulating the presentation of examination tasks and questions. It signifies the range of possibly relevant concepts and processes that might lead to a solution.

Marton, & Saljo (1976) introduced the parallel notion of **outcome space**, to signify the range of possible answers that pupils might give.

Problem space includes all of the knowledge a student has that might be relevant to the problem and its solution. This includes concepts, relationships, associations, strategies, and memories of past problems and how they were dealt with.

Outcome space considers all of the possible results that might be generated by students tackling a problem. It is difficult enough to predict all of the concepts that will be significantly activated in the minds of all the examination candidates; it will be much more difficult to anticipate all the conclusions they will come to. Yet this is exactly what we need to do in order to describe all the answers our markers will see. A perfect mark scheme is an evaluative description of the whole outcome space.

5 Language production

How do students compose a (written) answer?

Apart from a few exceptions, academic examinations require candidates to communicate their response to the examiner in writing. In very general terms the examining process can be described as mental measurement, or an attempt to measure some feature of the student's mind. Given how we have described learning, the aim is to measure the quality of the student's mental representation of the subject being examined. It follows, then, that a valid question will always ask for - in some sense - a summary of part of their understanding. (The question may, of course, ask for an application of knowledge, but it is not unreasonable still to see this as a summary for a particular purpose of the relevant understanding - linguists today insist that all forms of writing should be considered as writing for a particular purpose and audience.)

Summarising

The fundamental problem is this: if understanding is represented multi-modally in the brain, how does a student turn it into a textual summary? The mental model is not linear, and not linguistic, yet it has to be turned into a linear string of words. When the problem is expressed in this way, it does not seem surprising that writing a summary is a very difficult task.

Brown and Day (1983) report a study at different ages of the skills involved in summarising an artificially constructed text. Opportunities were created in the text for each of the five skills identified by van Dijk and Kintsch (1977) as involved in summarisation.

The table below shows the percentage of opportunities taken at each age (the column headings refer to the United States educational system, and indicate roughly 11, 13, 16 and 18 years of age):

		G5	G7	G10	Coll
1	Deletion : Trivial	95	93	91	95
2	Deletion : Redundant	96	95	93	98
3	Superordinate	52	51	82	85
4	Selection	28	33	52	53
5	Invention	14	23	38	46

The impression is that, for the more sophisticated aspects at least, summarisation is developmentally constrained: adolescents and even college age young adults find the process of writing in words a summary, *even of a verbal text*, intrinsically difficult. To construct a summary of a multi-modal representation of learning must be still more difficult.

Conclusion: Cognitive Psychology and Examining

Communication?

Examining is a process that depends on communication, usually linguistic. We often act as if the communication were unproblematic, but there are many opportunities for it to break down. Indeed, Spolsky (1993) suggests that we might do well to think of *misunderstanding* as more common than *understanding*. In our context, these failures may threaten the validity of the examination.

Consider what happens when a Marker, M, awards a mark to a response given by a Candidate, C, to a question set by a Setter, S, (this is the heart of the assessment process):

M evaluates
    M's interpretation of
        C's expression of
            C's answer to
                C's interpretation of
                    S's expression of
                        S's task, using
                            M's interpretation of
                                S's expression of
                                    S's demands.

All testing involves these three people, each trying to communicate with the others.

Some of the component judgements in this process refer to aspects of reliability; others, and notably those which complete the triangle by relating M back to S, refer to validity.

Summary

- All students' have an exam schema;
- Each student has only one mind, however many subjects they are studying;
- Their behaviour will be dominated by what they expect to happen;
- We must understand (and perhaps alter) their expectations;
- We must try to think how students think.

Generally, and especially at age 16 or less, they will have one fairly general schema covering exams, with specific components relating to the content and questions of each paper. Older students, more able to reflect on their behaviour and learning, may have more differentiated schemas for different exams.

They will have expectations about their feelings, and about the amount of time available. More importantly for question writers, they will have expectations about 'what will come up', what sorts of questions are asked in each area of study, what kinds of things they will and will not be expected to do. Under conditions of stress they will be less adaptable, and less able to deal with the unexpected, than normal.

When reviewing questions, we should try to think in the same way as typical young people, uncertain of their subject and feeling acutely stressed by the whole experience of a series of exams.

During our research, both in Edinburgh and in Cambridge, we have found many examples of unexpected responses which are understandable in terms of these theoretical perspectives - some where the student found a question unexpectedly easy, some where it was unexpectedly difficult (but mostly the latter!). Some examples will be given as illustration during the presentation of this paper.

References

Ahmed, A, & Pollitt, A, 1999. Curriculum demands and question difficulty. Paper presented at IAEA conference, Bled, Slovenia, 1999.

Anderson, R C, 1977. The notion of schemata and the educational enterprise. In RC Anderson, RJ Spiro and WE Montague, (eds) *Schooling and the Acquisition of Knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, R C, 1983. T*he Architecture of Cognition*. Cambridge, Mass: Harvard University Press.

Bachman, LF, (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bartlett, FC, 1932. *Remembering*. Cambridge: Cambridge University Press.

Brown, AL & Day, JD, 1983. Macrorules for summarizing texts: the development of expertise. *Journal of Verbal Learning and Verbal Behaviour*, **22**, 1-14.

Evans, J St BT, 1989. *Bias in Human Reasoning: Causes and Consequences*. Hove: Lawrence Erlbaum.

Garnham, A, 1987. *Mental Models as Representations of Discourse and Text. Chichester*: Ellis Horwood.

Gernsbacher, M A, (1990) *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum.

Grice, HP, 1975. Logic and conversation. In (eds) P Cole and J Morgan, *Syntax and Semantics 3: Speech Acts*. New York: Academic Press.

Hughes, S & Fisher-Hoch, H, 1997. Valid and invalid sources of difficulty in maths exam questions. Paper presented at IAEA conference, South Africa, 1997.

Johnson-Laird, PN, 1980. Mental models in cognitive science. *Cognitive Science*, **4**, 71-115.

Johnson-Laird, PN, 1981. Mental models of meaning. In (eds) AK Joshi, BL Webber and IA Sag, *Elements of Discourse Understanding*. Cambridge: Cambridge University Press.

McNamara, TF, (1996) *Measuring second language performance*. Addison Wesley Longman.

Marton, F & Saljo, R, 1976. On qualitative differences in learning: 1 - Outcome and Process. *British Journal of Educational Psychology*, **46**, 4-11.

Messick, S, 1988. Validity. In R. Linn (ed) *Educational Measurement*. New York: Macmillan.

Minsky, M, 1975. A framework for representing knowledge. In (ed) PH Winston, *The Psychology of Computer Vision*. New York: McGraw-Hill.

Pollitt, A, Entwistle, N , Hutchinson, C & de Luca, C, 1985. *What Makes Exam Questions Difficult?* Edinburgh: Scottish Academic Press.

Pollitt, A & Hutchinson, C, 1987. The validity of reading comprehension tests: What makes questions difficult. In D Vincent, AK Pugh & G Brooks (eds) *Assessing Reading*, Basingstoke: Macmillan Education.

Pollitt, A & Taylor, LB, 1996/7. Good tests aren't as good as they seem to be. Paper presented at LTRC conference, Jyväskylä, Finland, 1996. Reprinted in A Huhta, V Kohonen, L Kurki-Suonio & S Luomo, *Current Developments and Alternatives in Language Assessment*, University of Jyväskylä, 1997.

Sanford, AJ & Garrod, SC, 1981. *Understanding Written Language*. London: Wiley.

Schank, RC & Abelson, R, 1977. *Scripts, Plans, Goals and Understandings*. Hillsdale, NJ: Lawrence Erlbaum.

Spolsky, B, 1993. Comprehension Testing; or Can understanding be measured? In G Brown, K Malmkjaer, A Pollitt & J Williams (eds) *Language and Understanding*. Oxford: Oxford University Press.

Stenner, AJ, Smith, M & Burdick, DS, 1983. Towards a theory of construct definition. *Journal of Educational Measurement*, **20**, 305-312.

Sternberg, R, 1977. Component processes in analogical reasoning. *Psychological Review*, **84**, 353-378.

Taylor, LB, 1996. An Investigation Of Text-Removed Summary Completion As A Means Of Assessing Reading Comprehension Ability, Unpublished PhD thesis, University of Cambridge.

van Dijk, TA, 1981. Review of RO Freedle (ed) 1979. *Journal of Linguistics*, **17**, 140-148.

van Dijk, TA, & Kintsch, W, 1977. Cognitive psychology and discourse: Recalling and summarising stories. In WU Dressler (ed) *Trends in Text-Linguistics*. New York: De Gruyter.

Wason, PC & Evans, J St BT, 1975. Dual processes in reasoning? *Cognition*, **3**, 141-154.