

The question gets in the way: Inferring competence from examination performance

A symposium at the British Psychological Society London Conference, December 1999.

A collection of contributing papers by authors from the Research Division at the University of Cambridge Local Examinations Syndicate (now Cambridge Assessment)

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge.

Cambridge Assessment is a not-for-profit organisation.

© UCLES 2009

The question gets in the way: Inferring competence from examination performance

A symposium at the BPS London Conference, December 1999.

A collection of contributing papers by authors from the Research Division at the University of Cambridge Local Examinations Syndicate (now Cambridge Assessment)

Contents of this booklet

“The effects of stress on text comprehension and performance in examinations” by Dina Kiwan, Ayesha Ahmed and Alastair Pollitt	3
“Can Oral Examinations Improve the Psychological Validity of Assessment?” by Ayesha Ahmed, Alastair Pollitt and Leslie Rose.....	13
“The question gets in the way: inferring competence from examination performance; some closing remarks” by Jackie Greatorex	25

Additional Symposium Contributions

“Understanding and promoting better task design”
By Jim Ridgway , University of Durham and Tom Ormerod, Lancaster University

“Are examination standards all in the head? Experiments with examiners’ judgments of standards in A level examinations.”
By Jo-Anne Baird, AEB

Note that only the contributions from UCLES are in this booklet.

**“The effects of stress on text comprehension and performance in examinations” by
Dina Kiwan, Ayesha Ahmed and Alastair Pollitt**

University of Cambridge Local Examinations Syndicate

Paper presented at the BPS London Conference, December 1999.

Summary

This study aims to investigate the effects of time-induced stress on text comprehension, and the implications for performance in an examination setting. In the pilot phase, which has just been completed, we aimed to identify which features of language, context, or presentation, are more, or less sensitive to time-induced stress. We used a program called *Hypercard*, to present a narrative text on a series of computer screens to 4 participants, ranging in age from 10 years and 6 months to 11 year and 6 months. The participants were required to recall as much of the text as possible, followed by prompts to aid recall. Participants were then asked more specific questions to test their comprehension. Preliminary results from the pilot phase suggest that there are three key findings: i) participants had difficulty understanding who is narrating; ii) the use of schemas; and iii) participants had difficulty with the temporal sequence of the story.

Introduction

A range of linguistic, conceptual and contextual features have been identified as contributing to comprehension difficulty, (Brown, 1986; Brown, 1989; Brown, Sharkey & Brown, 1987). There have also been a number of studies that have been conducted to identify stressful conditions: “conditions under which linguistic and contextual factors interact so that the linguistic message is relatively easy or relatively more difficult to understand” (Brown & Markman, 1991). Examinations are a source of stress in various ways, and there is considerable evidence in the literature that stress affects task performance, often negatively. One simple form of stress, which is relatively easy to manipulate (and relevant to the examination context) is time. This study aims to investigate the effects of time-induced stress on the process of pupils reading a text in an examination setting, and the resulting effects on their comprehension of this text.

A pupil’s ability to understand a text and the questions relating to it will inevitably affect their performance on these questions. The reader has to carry out processing at several levels – at the individual word level, at the sentence level, at the groups-of –sentences level, and at the text level (Gerrig, 1986). Reading a text and answering questions may be difficult for a number of reasons. The accessibility of the linguistic, contextual, and conceptual features in the text may cause difficulties in comprehension (Bell, 1995). For example, there may be unfamiliar words, complex syntax, temporal complexity, unclear or delayed identification of the referent, and the causes and/or intentions in the text may not be explicit, requiring the reader to be able to make inferences. There is a distinction, however, between the pupil’s ability to read a text with intended, and therefore valid ‘sources of difficulty’ in the task (Pollitt & Hutchinson, 1985), and invalid ‘sources of difficulty’. These are aspects of the text and/or questions that impede assessment of what the questions are intended to test. The primary function of a question is to instruct the pupil what to do; it should not itself be a source of serious difficulty.

In addition to linguistic, contextual and conceptual features that may cause difficulties in text comprehension, there is a sizeable body of research on the role of working memory in

reading. It has been proposed that those readers who have difficulty comprehending a text may be less efficient at storing and processing information at the same time in the working memory system. As a consequence, it will be relatively more difficult for these readers to form a coherent model of the text (Oakhill and Cain, 1997). There have also been a number of studies suggesting that there is a strong relationship between comprehension ability and working memory capacity, but a causal link has not been established. (In fact, differential reading experience may lead to the relationship between working memory capacity and comprehension ability.)

Incomplete knowledge may also pose a problem for readers in their attempts to comprehend a text. In order to make inferences, the reader has to be able to use knowledge stored in memory, referred to as a *schema* (Bartlett, 1932). Schema theory proposes that information is gathered together into meaningful units, and that we have schemas for most things that we have experienced or encountered. A script is a memory structure that is composed of a sequence of actions that occur in a given stereotypical situation (Abelson, 1981). We were also interested in exploring the effects of stress on these processes, and we hypothesised that, under stress, pupils would be particularly likely to rely on commonly-held schema in their comprehension of the text.

Given that assessment methods are rapidly developing, and that computerised assessment techniques are becoming ever more widespread, we decided to explore the effects of stress on comprehension by presenting the text on a computer screen at different speeds. Computers are now widely used both at school and at home in the development of reading, with a variety of programs, (e.g. programs which provide practice for improving speed and fluency of word recognition, or practice for learning phonic skills) (Singleton, 1997). However, there has, as yet been relatively little research on using the computer for the assessment of reading. One study conducted by the National Council for Educational Technology (NCET) in 1995 compared a computerised test of reading comprehension with a conventional paper-based assessment. The main conclusions of this study were that with the computerised version of the test, the pupils (aged 7-10 years) were more motivated, and that there was greater objectivity of assessment.

A second study (O'Hara & Sellen, 1997) compared reading paper and on-line documents, focussing on the reading processes for the purposes of producing a written summary. They concluded that for such a task, reading on paper has the advantages of allowing the reader to annotate while reading, and to cross-refer between pages, thus facilitating a deeper structural understanding of the text.

We have just completed the pilot phase of this study. We aim to investigate the differential effects of stress using a program called *Hypercard* which presents the text on a series of computer screens. We hypothesised that the process by which pupils read and understand texts are particularly sensitive to time-induced stress. We aim to identify which features of language, context, or presentation, are more, or less sensitive to time-induced stress.

Method

In this pilot phase, our aim was mainly exploratory, in which we attempted to:

- i determine three appropriate speeds (a 'low', a 'medium' and a 'high' speed) at which to set the computerised delivery of text.
- ii evaluate the suitability of the text for Year 6 pupils (10-11 years of age).
- iii identify potential linguistic, contextual, conceptual or presentational features that may contribute to comprehension difficulties under time stress.

Participants included 4 subjects (3 boys and 1 girl), ranging in age from 10 years and 6 months of age and 11 years and 6 months of age. Each participant was presented initially with three practice texts, to enable them to familiarise themselves with the computer

paradigm, and to reduce performance anxiety. The software was programmed so that each practice text was presented at a different speed (2.5 words, 3.0, and 3.5 words/second), with each screen presenting one paragraph of the text at a time. The rate of presentation and the texts was counterbalanced across participants. After each practice text had been presented, the subjects were asked if they had been able to read all the text on each screen. This, in effect, was an informal calibration, allowing us to make an approximate assessment of the participants' optimal reading speed. Participants were then presented with the actual text, which was delivered at a rate that provided a moderate challenge (i.e. stress) to each subject. Participants were told that the actual text was longer than the practice texts.

For all texts, the title appears in the centre of the screen, and when the participant indicates that he/she is ready to start, the experimenter clicks on the mouse to bring up the short introduction that precedes each text. Again, participants can read this in their own time, and when they are ready to begin, the experimenter clicks on the mouse again to start the program. Each paragraph of the text appears on the screen, with a visual indicator at the bottom of the screen so that participants can monitor the time available to read what is on the screen. This visual indicator is in the form of a horizontal black bar that gradually turns white. When it has turned completely white, this means that there is no reading time left and the next screen automatically appears. This process repeats itself until the end of the text is reached.

This procedure was explained to each participant in the following manner:

"Now you're going to read a story on the computer. When we start, you will see the title of the story. When you're ready, the introduction will come on the screen. Then when you're ready, I'll start the story. (Before you read the actual story, we'll do a few practice ones together so that you get used to it.) When the story starts, you have to read what is on the screen before the next screen comes on. There is a black bar at the bottom of the screen that gradually turns white. When it turns all white, this means that your reading time is finished on this screen and the next screen will come on. You will have to read fairly quickly but you should still try and read everything properly, as we will be asking you about the story at the end. You don't have to press any buttons on the computer. The next screen will come on by itself."

The presentation of each practice text lasted approximately 1 minute. After each practice text, each participant was asked:

"How was it? Could you read everything on the screen?"

The actual text that participants were tested on is as follows:

The Story¹

Introduction: This story is about John and his family....

We arrived back home just after John. We saw him park by the back door, take something out of his saddle-bag and run upstairs in front of us with it. He went straight into the bathroom. As John ran up the stairs, Mum called after him, asking him if he'd enjoyed his ride. He answered that it'd been great, and that he'd cycled forty-three miles. He shut the bathroom door behind him.

It was his turn to help to get the tea ready. I stayed in my room doing some revision. Dad called upstairs to ask me to get a clean table-cloth from the airing cupboard in the bathroom. John immediately bolted up the stairs, saying he'd get it. But I was already at the cupboard. I opened the door and something came at me in a fury, beating against my face. I screamed out and John pulled me to one side.

¹ Adapted extract from 'Granny was a Buffer Girl' by Berlie Doherty.

Mum followed him in, demanding to know what was going on. John was holding a grey pigeon. It fluttered frantically in his hands, its eyes bright with terror. I was still shaking. I'll never forget the thrust and flap of those wings across my face. Mum started to pull out dirty linen from the airing cupboard, asking how a bird had got in there. John replied that he'd put it there, and apologised, explaining that he and his friend Harry had found it on their ride. He explained that Harry had thought it was dead, but that he didn't, so he brought it home in his saddle-bag, and then put it in the airing cupboard to get warm.

Mum, who was still as upset as I was, shouted at John, calling him an idiot, and complained about the mess the bird had made in the airing cupboard. She told him to get it out. John just stood there with his hands clasped round the pigeon, holding it against his chest so that it looked strangely like a beating heart. Again Mum shouted that he should get it out, adding that she couldn't stand birds in the house. Then Dad, still floury from the pastry he'd been making with John, came in to see what was happening. He told John to get rid of it.

John became defensive asking, "What d'you mean, get rid of it?"

Dad told him to take it back to where he'd got it from. When John said he'd found it near Grindleford, which was miles away, Dad still insisted he get rid of it. John said he wanted his tea first. But Dad was firm, saying his tea could wait since he had to get rid of it first.

John's face looked pinched and white. I was the cause of the trouble with my screaming, and I could have helped him to get out of it. But I didn't. I was pleased that they had sided with me against him, protecting me from my fright. I told him that it was a stupid trick to play on me. He argued that it wasn't a trick, since he'd put it there to get warm. But Dad just steered him out onto the landing, telling him to clear off with it. I smiled at John as he went past, smug with victory.

He came into the room and flopped into an armchair, sprawling out his arms and legs. He was exhausted. His face was streaked with dirt where he'd been rubbing his eyes. He calmly stated that he'd brought it back. Dad came in from the kitchen and stood in the doorway, too angry to speak. John gently undid his jacket and brought out the bird that was nestling there. He cradled it in his hands and it watched us quietly. Paddy, the cat arched his back, and bushed his tail out at it. I stroked him while he hissed.

Mum asked why he'd brought it back. John then told us how he'd taken it all the way back to where he found it, how he put it back in the trees in the Longshaw estate, and how it began to settle down spreading its wings out as he walked away. But then he decided to go back to see if it was alright. When he found it in the same place, and still in the same position, he picked it up and tossed it out to make it fly. But because it just dropped, John thought he'd killed it. He informed us that since it couldn't fly, he'd brought it home to look after it.

I told him that it wasn't his responsibility, but he retorted that it was, saying, "I found it. You don't expect me to just leave it there, do you, when it can't look after itself!"

I caught the quick look that passed between Mum and Dad, and couldn't fathom it. Mum pushed her books away wearily, and turned to Dad, "Give John his dinner. I think there's an empty carton in the car boot. I'll fetch it."

After participants had finished reading the actual text (3-5 minutes), a delay was introduced to stabilise the memory of the text. This filler task*, shown below, consists of 10 test items, in which the participants were required to 'circle the odd one out' out of four options, and explain their choice orally. As well as being a distractor activity, this task required semantic processing. This activity took approximately 2-3 minutes, and was administered by a second experimenter, whilst the first experimenter prepared materials and the tape-recorder for the recall and comprehension phase relating to the text.

***Circle the odd one out.**

Example: *bracelet* *ring* *necklace* **pen**
Pen is different. The others are jewellery.

- | | | | | |
|-----|---------|-----------|---------|-----------|
| 1. | tea | coffee | bread | milk |
| 2. | kitchen | bathroom | bedroom | garage |
| 3. | green | big | blue | red |
| 4. | fork | spoon | knife | cup |
| 5. | water | meat | bread | fish |
| 6. | car | sheep | train | bicycle |
| 7. | Monday | Thursday | Sunday | Tuesday |
| 8. | July | Christmas | March | January |
| 9. | book | letter | TV | newspaper |
| 10. | uncle | friend | sister | mother |

After participants had completed the filler task, they were told that they would be asked some questions on the story:

"Now I'm going to ask you about the story and I'm going to tape it because I won't remember everything and so that I don't have to make notes:"

A microphone was then clipped to the participant, and the following questions were asked:

(General Comments – "how was it?")

A/ Free Recall

Can you retell the story to me, in as much detail as you can remember?

Before, starting with the prompts and the questions, participants were told that they would be going through the story again in order to extract a few more details. This was to provide a rationale for what may have been perceived by participants to be a repetitive exercise:

"Now I'm going to ask you for a few more details so it may feel like you're repeating yourself, but let's just go through it to try and remember as much as you can."

B/ Prompts

Who do you think is telling the story?

Can you remember any more details about John arriving home at the start of the story?

Can you remember anything more about what happened when Dad asked for a table-cloth?

Can you remember anything more about how John's brother/sister felt (when he found the bird)?

Can you remember anything more about how Mum felt (when she saw the bird?)

Why did John bring the bird home?

Can you remember anything more about what John was told to do?

How do you think John felt about this?

And what about John's brother/sister – how did he/she feel?

Can you remember anything more about what John did next?

Can you remember anything more about what happened at the end of the story?

C/ Questions

1.1 At the start of the story, who arrived home first?

Why did John put the bird in the airing cupboard?

What did John have to do before his tea?

What did John do when his Dad told him to get rid of the bird?

(Why did John bring the bird back the second time?)

How did John's Dad feel when John brought the bird back the second time?)

At the end of the story, what do you think happens next? Why?

Participants were initially asked to recall as much as possible, followed by prompts to aid recall. Then participants were asked more specific questions to test their comprehension. On completion of the above (a further 10-15 minutes), each participant was given a Mars bar as a token of appreciation for taking part in the study.

Results and Discussion

Participants were able to familiarise themselves with the paradigm of the computer presentation using the practice texts. These practice texts were typically presented at three different speeds: 2.5, 3.0, and 3.5 words/second. According to their responses to the question, "Did you manage to read everything on the screen?", we were able to determine an appropriate presentation speed that would challenge the participant. One participant was presented the actual text at a speed of : 2.5 words/second, two participants were presented it at 3.0 words/second and one participant at 3.5words/second.

All four participants satisfactorily completed the filler task with little difficulty, although one participant required prompting on two of the items.

The taped recordings of the responses of all the participants on the recall task, followed by the prompts and questions were transcribed.² Several key issues emerged from a qualitative analysis of these transcriptions:

- i) difficulty understanding who is narrating
- ii) the use of schemas
- iii) difficulty with the temporal sequence of the story

i) Two out of the four participants had difficulty identifying the narrator of the text, in response to the direct question, "Who do you think is telling the story?" One participant was reasonably able to recall the story, which he did from the viewpoint of an observer. Yet in response to the above question, he became confused, saying "...is it...is it him? No, his Mum first...I think sometimes it's him and sometimes it's his Mum." The second participant spontaneously provided the term, 'narrator', yet when prompted, he tentatively suggested that the narrator was 'John', the main character. It may be that, under conditions of stress, simplification occurs, with recall of only the key characters, and no new characters are introduced. Indeed, the first participant mentioned above stated that there were three characters in the story – 'John', 'Mum' and 'Dad', and the second participant mentioned only two characters – 'John' and 'Mum'. The other two participants, a boy and a girl, both make the inference that the narrator is John's brother, even though there is no explicit reference in the text that the sibling is a brother rather than a sister.

ii) The use of schemas is particularly evident with two of the participants. Firstly, typical schemas relating to the roles of 'Mum' and 'Dad' appear to have had an influence on the recall and understanding of the text. One participant recalled that "Mum went down to make dinner", even though there is no explicit mention of what 'Mum' is actually doing, and in fact, the text explicitly states that it is 'Dad' who is getting the tea ready. In addition, a second participant recalls that 'Dad' plays the assertive, decision-making role telling John to get rid of the bird, even though in the text, 'Mum' tells him to get it out before 'Dad' does. It is also interesting to note that 'Mum' has a more prominent role than 'Dad' in the retelling of the story. In fact, in one participant's recall there is no mention of 'Dad' at all.

A second schema that may be coming into play could be a schema relating to sibling rivalry. Two of the participants suggest that the narrator is a brother rather than a sister. It could be that they are using a schema relating to sibling rivalry that may be stronger between brothers, rather than between a brother and a sister.

Thirdly, it is interesting to note that two of the participants recall that 'John' locks the bathroom door, even though this is not mentioned in the text. This could be due to the participants having a schema for 'going to the bathroom' that includes locking the door. However, it seems, that at least for one of these participants, the word 'bolted', used a few sentences

² See Appendix for transcriptions for the four participants.

later, in a different context – ‘John *bolted* up the stairs’, has been incorporated into the schema of ‘going to the bathroom’: In response to the question, “..What happened when Dad asked for a tablecloth?”, the participant responded “Um...John bolted ...bolted up the, um, the door...so John’s brother had to do it.”

Finally, there may be some evidence for participants exhibiting a simplified understanding of the emotions expressed by the characters in the text. They may be using very broad schemas of ‘positive’ versus ‘negative’ types of emotion. The participants appear to be aware that there are mostly negative emotions being portrayed in the text, but seem unable to clearly distinguish between them. For example, two participants note only that ‘Mum’ was ‘annoyed’ or ‘angry’, but do not seem to be able to comment on the emotions exhibited by the other characters. In addition, one participant tentatively suggested that John’s brother was ‘angry’, and ‘Mum’ was ‘scared’ on finding the bird; and in response to the question “How did John’s Dad feel when he brought it back a second time?”, he at first said that he couldn’t remember, and then again tentatively suggested “I think he was angry.”

This evidence suggests that when under stress, pupils are more likely to utilise a pre-existing memory structure that incorporates stereotypical information relevant to comprehension of a particular concept, role, or situation. This schematic information tends to be stereotypical as it has been abstracted over a number of different occasions (Gerrig, 1986). Participants may be utilising these automatic schemas that are not modified due to the time stress, and also because they may only be deriving partial information from the text due to lack of time.

iii) This text is complex in terms of its temporal sequencing of events. Towards the end of the text the third paragraph before the end is out of temporal chronological sequence. This paragraph recounts events that actually occur before the events recounted in the paragraph starting ‘He came into the room...’. In addition, there is a significant jump in time between the end of the preceding paragraph ending in ‘...smug with victory’ and this paragraph starting ‘He came into the room...’.

Only one participant was able to recount a relatively coherent story incorporating these difficult temporal aspects of the text. The other three participants did not realise that ‘John’ had in fact taken the bird back and returned before he had his tea. Participants were not prompted relating to these events towards the end of the story if they did not exhibit an understanding of the events in their initial recall or during the course of being prompted.

Given that the computer program presents the story in paragraphs, with approximately one paragraph/screen, it could be that this presentation may have made the task of forming a coherent overview of the temporal sequencing in the text more difficult, and especially under time stress. It could, however, be that this text is just too difficult for 10-11 year olds. The cues in this paragraph are very subtle in the form of the past perfect tense ‘he’d’. In fact, some of the verbs in this paragraph are not always in the past perfect tense, simply for stylistic reasons – that repeated use of the past perfect would read awkwardly. We intend to investigate these possibilities in the next phase of the study by introducing a number of control conditions: the first condition would be the ‘non-stress’ condition – participants themselves would be able to control the rate of presentation of the text on the screen; the second condition would be the ‘paper’ condition, where a break in the text could be marked by an increased space, and participants would be able refer back through the text if they needed to, and the third condition could be a ‘scroll-down’ condition, in which the whole text is presented on the screen, which would scroll down automatically. Yet another condition - the ‘direct speech’ condition could include the original version of the above text which is predominantly composed of direct speech. It could be that this version may cue in the temporal aspects of the text to a greater extent than our adapted version that we have used in this pilot. Alternatively, this text could be trialled on an older cohort of pupils – aged 13-14 years.

Implications for the next phase:

In addition to the above-mentioned controls that would be introduced, we propose to introduce some further prompts. These prompts would be more precise, focussing on specific details. This is because there were many details in the text that the subjects omitted, and it was not possible to determine what had been forgotten or misunderstood.

Possible prompts could be:

Do you remember anything more about:

- 'the table-cloth'
- 'flour/pastry'
- 'flapping'
- 'a trick'
- 'Harry'
- 'smug with victory'

What do you think this means:

- *'I caught the quick look that passed between Mum and Dad, and couldn't fathom it. Mum pushed her books away wearily, and turned to Dad, "Give John his dinner. I think there's an empty carton in the car boot. I'll fetch it."*

Given that this study is still at the pilot phase, conclusions must be very tentative. The strongest finding appears to be that, when under stress, pupils are particularly likely to use schemas to structure their understanding and recall, even when these do not match with explicit references in the text. One could venture to raise the controversial question of whether introducing non-stereotypical situations within texts in examination conditions (i.e. under conditions of stress) is actually introducing an invalid 'source of difficulty'. Is it appropriate to use texts with non-stereotypical situations or characters with non-stereotypical roles in examinations? Or should such texts be reserved for use only in teaching and learning settings? If, for example, the aim of an English exam is to assess certain skills, such as retrieval of information, inference, synthesis, comment on language, etc, then the text is the vehicle through which these skills are assessed and should not introduce further, unrelated invalid sources of difficulty.

This can apply to questions in other disciplines. This issue is exemplified by a question from a 1999 GCSE paper in Chemistry:

Put a ring around the name of an element which is in the same period as sulphur

*fluorine
magnesium
oxygen
potassium*

The correct answer is *magnesium*, yet most students gave the answer, *oxygen*, even though they had access to the periodic table. Students tended to make this error because questions relating to the periodic table usually refer the *group*, rather than the *period* that an element is in. In fact, this question initially misled a marker marking this question.

This is because the word, *period*, was unexpected, a less familiar concept.

It should be noted however, that it is a difficult task to always determine what a 'stereotypical' situation or role may be, and hence difficult to clearly determine which text or question may contain 'non-stereotypical' situations, roles or concepts, and hence 'invalid' sources of difficulty.

References:

Abelson, R. P.(1981). Psychological status of the script concept. *American Psychologist*, **36**, 715-29.

Bartlett, F. C.(1932). *Remembering*. Cambridge: Cambridge University Press.

Bell, J. (1995). *What makes language difficult to understand? UCLES notes*.

Brown, G. (1986). Investigating listening comprehension in context. *Applied Linguistics*, **7**, 284-302.

Brown, G. (1989). Making sense: The interaction of linguistic expression and contextual information. *Applied Linguistics*, **10**, 97-109.

Brown, G., & Markman, S. (1991). Discourse processing and preferred information. *Linguistics and Education*, **3**, 47-62.

Brown, G. D., Sharkey, A., & Brown, G. (1987). Factors affecting success of referential communication. *Journal of Psycholinguistic Research*, **16**, 535-549.

Gerrig, R (1986). Text Comprehension in *The Psychology of Human Thought*. R. Sternberg, and E. Smith (Eds.) Cambridge: Cambridge University Press.

Oakhill, J., & Cain, K. (1997). Assessment of comprehension in reading. In *The Psychological Assessment of Reading*. J.R. Beech & C. Singleton (Eds.). NY: Routledge.

O'Hara, K., & Sellen, A. (1995). A Comparison of Reading Paper and On-Line Documents. *Proceedings of CHI 1997 conference, Atlanta, GA*.

Pollitt, A. & Hutchinson, C. (1985). The Validity of Reading Comprehension Tests: What Makes Questions Difficult? In *Assessing Reading*, Proceedings of the UKRA Colloquium on the testing and Assessment of Reading.

Singleton, C. (1997). Computer-based assessment of reading. In *The Psychological Assessment of Reading*. J.R. Beech & C. Singleton. (Eds.) NY: Routledge

**“Can Oral Examinations Improve the Psychological Validity of Assessment?” by
Ayesha Ahmed, Alastair Pollitt and Leslie Rose**

University of Cambridge Local Examinations Syndicate

Paper presented at the BPS London Conference, December 1999.

Abstract

Oral assessment has often been disregarded due to problems of reliability. However, it may provide a more valid way of assessing candidates' level of knowledge than the widespread method of written assessment. Oral assessment offers the opportunity for the marker to enter into a dialogue with the candidate, and to use strategies such as prompting which can improve communication with the candidate. It also involves different time constraints from written examinations. At present, other than in Modern Languages, oral tests are used only in Certificate of Achievement (CoA) examinations, but they may be appropriate at other levels.

Questionnaires were sent to teachers carrying out the oral examination for Geography CoA, and 4 teachers and 16 students were interviewed. Transcripts of 18 of the teachers' orals have been analysed in detail to discover more about the language used, and the relationship between oral and written grades has also been investigated.

Introduction

Oral exams have historically been a popular method for assessment, but in the last 50 years the majority of assessment in Britain has been in the form of written tests, largely due to criticisms of the reliability of orals (e.g. Hartog and Rhodes, 1936). However, the proportion of the population being formally assessed at age 16 has increased dramatically in that period from the top 10% of the ability range to around 99% of students. As we are now assessing such a wide ability range at age 16 we need to consider a variety of methods for assessment in order to ensure that we are assessing all students in the most appropriate way. For those students who are low achievers the written test may not be the ideal tool with which to assess competence. These students have difficulty with writing and spelling that might prevent them from demonstrating their knowledge in a written exam. For these students oral exams provide a possible alternative, and this paper considers some of the issues this possibility raises.

Assessing the ability to talk about a topic rather than the ability to write about it eliminates the need for the student to develop good written communication skills. In a written exam we cannot help testing the student's ability to write – to turn thoughts into a linear string of words (Pollitt & Ahmed, 1999). In some cases it is valid to test this ability as we want students to learn how to write, but in other cases this gets in the way of finding out what they have learned about the subject. It can result in an assessment of performance on a writing task rather than competence at a particular subject. One of the purposes of oral exams is to provide a way of assessing students' knowledge rather than their written communication skills. Oral exams are still widely used in assessing Modern Foreign Languages, but these orals have a different purpose, which is to test students' verbal communication skills.

These two aims, assessing knowledge of a subject and assessing verbal communication skills, coincide in foreign language orals but may be in conflict in an oral exam for any other subject. For example, oral exams in Geography Certificate of Achievement (CoA) are designed to allow low-achieving students the maximum chance for positive achievement in Geography, and the examiners have chosen to do this through the medium of oral communication. It is hoped this would allow examiners to form a clearer idea of students'

knowledge of Geography than they would in a written exam, as these students cannot write well and lose confidence in written exams. However, for some students this method may backfire as they have little confidence in an oral exam setting and feel more at ease when writing as they do not have to 'think on their feet'. Oral communication skills are still developing at age 16 (Dickson, 1981) and there are also large individual differences at this age.

If the aim of the oral in Geography is to help students by avoiding writing problems then it succeeds for some, but not for others for whom talking about the subject is just as challenging, if not more challenging. However, there is another advantage of oral exams, and that is that they are interactive. They involve a dialogue between student and assessor. The assessors (usually teachers) have the opportunity to guide the students through the questions, to prompt them and to encourage them. This may allow the students to leave the exam with a feeling of positive achievement that is not necessarily related to the number of marks they have obtained. It also changes the nature of the communication between the student and the examiner, and should result in fewer misunderstandings of questions.

Oral Examining in Geography Certificate of Achievement

The Geography Certificate of Achievement (CoA) is an exam designed for students aged 16 who are not likely to be able to achieve a grade in GCSE Geography. The CoA is currently available in Basic Literacy, IT, Maths, History, Geography, Design & Technology, Modern Languages, Physical Education, Religious Education, Science and Building Studies. There is an oral exam in the Languages (including Basic Literacy), and in some (but not all) syllabuses in Mathematics and Geography. This paper considers the Geography 'Syllabus C' Certificate of Achievement from OCR, which consists of a written paper, written coursework, and the oral interview. There were 630 entries to this exam in 1999, around 2.5% of the total entry for GCSE Geography 'Syllabus C'. The oral interview is designed to last approximately 10 minutes and is on a different set topic each year. The students are given a Resource Booklet to look at in their lessons in the weeks preceding the exam. During the exam *their own teacher* asks each individual student some set questions about the resources, and the student answers orally. Each session is recorded on audio tape. These orals represent the first re-introduction of oral methods into public examining in UK schools (other than language exams) for 50 years, and are intended particularly for low-achieving students who used not to be assessed formally at all.

A major issue for this exam is the way in which the teachers ask the questions, and the verbal prompts that they give to the students. The teachers are told that they may be flexible with the way in which they ask the questions, and vary the exact wording to suit their own interviewing style. The teachers are also told that they can ask supplementary questions so that students can achieve something positive. Positive achievement is seen by the examiners as crucially important for CoA candidates, and this is aimed for sometimes at the expense of reliability. In order to give these students a sense of positive achievement, many of the teachers (but not all) are guiding students' answers by giving prompts when they answer incorrectly or do not answer at all. The teachers use a variety of approaches when prompting the students. These include requesting further information; rephrasing the question into a more structured form; giving the student extra information; or simply repeating the original question.

The potential for variability in this process is obvious, and would need to be controlled, partly by training but also through a process of moderating the results that different teachers produce. This is, of course, the basis for Hartog & Rhodes' challenge sixty years ago, and would need to be addressed seriously if oral examinations were to be given more prominence again. This paper does not attempt to deal with the problem of moderation, but does bear on issues of training, and analyses the benefits that we might hope to see if satisfactory checks could be developed.

Data

Questionnaires were sent to all of the 87 schools entering pupils for Syllabus C exams in 1999, and 36 completed questionnaires were returned. These gave the teacher's views on aspects of the oral such as issues of prompting and marking. Some of the results are cited below. Four teachers and sixteen students were interviewed in detail about their experiences in the oral exam and their views can also be seen below. A detailed analysis of twelve transcripts of taped orals was carried out, along with a comparison of marks on the oral exam, the written coursework and the written exam.

Prompting

Utley, Mitchell, & Phillips (1983) conducted a review of oral tests and concluded that they can be suitable for subjects other than languages. They suggested that it is necessary to have a clear policy for prompting in order to ensure reliability. The policy can vary from allowing no prompts at all to allowing teachers to repeat or re-phrase questions, or even leaving the prompting to the teacher's discretion, which is what occurred in the Geography oral. There were clear differences in the ways in which different teachers decided to prompt their students, and in the ways in which one teacher prompted different students. Utley et. al. (1983) suggest that one possible solution lies in writing questions that are specific enough or transparent enough to need no prompting. However, starting with very specific questions can lead to different problems of prompting. If the student cannot answer the very specific question what does the teacher do? There is a choice of either moving on to the next question or trying to re-phrase the question. In this case the re-phrasing cannot consist of making the question more specific, so instead the teacher asks a slightly different question. This can lead to ambiguity about what the task actually is.

Another important issue is when to prompt the student. Should teachers prompt after a certain length of silence, or should they prompt if the wrong answer is given? Should they prompt until they get the right answer or until they decide they will never get the right answer? Should they prompt until the student feels a sense of positive achievement? Finally, how should the amount and nature of prompting be reflected in the marking?

It is impossible to give instructions for prompting that can be applied by every teacher to every student. Each student needs a different amount of support. The issue here is the sort of model we are following for assessment. Assessment can be conducted by gradually adding higher hurdles to an easy task, or alternatively by gradually removing support from the student to see if they can still complete the task. The number of hurdles passed or the amount of support needed can be used as an indication of the level of the student. We run into difficulty when we try to combine these two models for assessment. In the case of the Geography CoA oral, the teachers are giving different levels of support to the students, but the students are then being assessed on their performance on the task, that is the number of hurdles they achieve in the exam. They are assessed, using a generic mark scheme, on the amount of information they recall, the number of terms they can define and use, whether they can identify and give explanations for patterns and relationships and so on.

It is possible to remove support gradually from a series of written questions, but for individual questions in the oral it is more appropriate to add support gradually, which is what the teachers are doing when they are prompting. If this is the desired technique for the oral then it must be reflected in the mark scheme, with more credit gained for achievements without support and less if support was needed. However, problems would occur if the mark scheme consisted of a combination of statements about levels of performance on a task and statements about levels of support needed. Without support, students doing the CoA may not be able to identify the opportunities in a question, and may only give minimal responses.

Adding support gives every student a chance to exploit the opportunities in a question.

The questionnaire data from the teachers who carried out the oral assessment in 1999 showed that 91% of the teachers had to re-phrase questions, with 88% giving prompts to help the students answer. Most of the teachers felt that they had been consistent in the degree of support they gave each student, although 68% said they had to change their approach to give each student the best chance of answering the question. The comments the teachers made are more revealing. Some of the teachers were not confident about how to prompt and how much to prompt:

‘I tried to be flexible with the questions but found it difficult.’

‘Prompting is hard to do without ‘helping’ the candidate.’

‘I needed to lead the candidates to get more detailed answers but I wasn’t sure how much was allowable.’

‘...but how far can you go to lead them? How far should you coax them?’

Others took a slightly different view:

‘...even if I had prompted too much, and given them too much help, the most important thing was for them to go out of the room with a good feeling rather than going out feeling they’d failed...’

In a detailed analysis of transcripts of 12 of the teachers, 10 of them prompted with a request for more information and 11 of them used prompts that directed students in how to use the resource to answer the questions. For example:

Teacher: ‘Which part of the world has the largest area of rain forest at present?’

Student: ‘Um. Asia.’

Teacher: ‘Have another look at the map. Just use the map.’

Student: ‘Latin America.’

The teachers were not consistent unless they were giving no prompts at all. Although 10 of them prompted for more information, this was only done for 25 out of the 43 students involved.

Sixty-eight percent of the teachers said they felt confident about the marks they then gave, although some clearly felt differently:

‘...we are not experts at assessment.’

‘I just wasn’t happy with the marks I gave. You just can’t get the level - I feel dubious about my marking...’

Seventy-four percent of teachers said that they tried to take into account the amount of prompting they had given when marking the orals. One teacher said:

‘They would get higher marks if they had less help.’

Assessment by Teachers

Most of the teachers had no training in how to use a generic mark scheme, and were trying to combine its use to assess the task with their knowledge of the level of support they had given. Despite difficulties with the assessment and a lack of confidence in doing it, 91% of teachers said they wanted to carry on doing the assessment for the oral themselves and not have it conducted by external examiners. The students also said they preferred to be assessed by their own teachers.

Teachers have the advantage of knowing the student they are assessing. They know how

much the student knows about the subject based on their performance throughout the course, so they know what to expect in the exam. If the student is unable to answer a question in the oral and the teacher knows that they should know the answer, the teacher can give prompts. An external examiner would not have the advantage of trying to elicit information that they know the student has. However, along with knowing the student comes a lack of objectivity. Teachers may assess students' performance based on their whole knowledge of that student, not on the student's performance in the oral alone. Can the teacher really separate their knowledge of what the student *could have* demonstrated in the exam from what they *actually* demonstrated during that ten minute period? What does the teacher do who knows that the student knows the answer but cannot get them to say it? There is a temptation to keep prompting until the right answer is reached, or to mark as if the right answer was reached. If we are aiming to assess students' overall ability then we need not worry about teachers taking into account performance throughout the course, as long as all teachers do this consistently. However, if we are assessing the student's performance on the particular task then judgements must be made on the task alone. We have therefore to decide whether we want our measurement instrument to be the teacher or the task.

Cognitive Demands and Affect

Attention

There may be different cognitive processes occurring in students' minds when they are answering questions in oral and written exams. Apart from the lack of a need to translate thoughts into a string of written words, there are also different demands made on students in oral exams. One of these is associated with the attentional processes needed. In an oral exam the assessor is directing the student to the task by asking the question. Students do not have to control their own attention (as they would on a written paper) because it is externally directed by the assessor, who can direct them explicitly to the resources but also more implicitly, using prompts, to the ideas they should be using. This results in what seems like less conscious effort for the student. As some of the students put it:

'You just answer the question and then go – it's easier than normal.'

'It didn't feel like an exam – there was no pressure – you just sit there.'

Because the students' attention is directed by the teacher, and because the teacher gives prompts, some students are able to achieve more than they would in a written test. As the teachers put it:

'Many of our candidates 'shone' in comparison to their written work.'

'The oral assessment gave candidates an opportunity to show what they knew only when I directed the questions in that direction- I had to squeeze it out of them.'

The problems with the oral assessment from the students' points of view were mostly to do with affective factors such as feeling nervous. Seventy-four percent of the teachers thought that most of their candidates preferred the oral assessment to written tests. For example:

'They like oral exams as CoA candidates tend to struggle with writing so are able to get a sense of achievement.'

'Some felt nervous but afterwards generally they were pleased to have demonstrated knowledge and understanding.'

'Whilst they found it hard, I feel that written assessments are harder for them. We need to embed oral work in our Scheme of Work to increase student familiarity.'

The last quotation raises an important point which is that students do well in assessments with which they are familiar. For a familiar task they may be able to apply a schema, i.e. a mental representation of how to tackle that sort of problem, but if they have not done an oral before

they may not have a schema for coping with it. This illustrates the importance of having practice runs and mock orals.

Despite the differences discussed so far between written and oral assessments, a significant positive correlation ($r = 0.53$) was found between performance on the written coursework task and performance on the oral exam in an analysis of the results of 60 of the students. This does not necessarily indicate that the two tasks are assessing the same kinds of skills, but simply that those who did well on one tended also to do well on the other. However, the correlation between the oral and the written exam was only 0.40. Those who did well in the oral tended to also do well in the written exam but the group who did less well on the oral achieved a wide range of scores in the written exam. Those students obtaining low scores on the oral but high scores in the written exam may have been suffering from a lack of confidence in the oral exam.

Time-Management

Another factor that differs between written and oral exams is the amount of time students have to answer the questions. The time pressures in an oral are quite different from those in a written paper. There are usually time guidelines for an oral but no strict rule. This takes away the pressure from the student of managing time and spending the correct amount of time on each question or group of questions. However, there are different time-related pressures in an oral exam. The student has to answer questions on the spot, without time to stop and think for longer than the teacher allows, and often without a chance to leave a question they cannot do and then come back to it later. The time is managed not by the student in an oral but by the teacher. It is the teacher who decides when the student is ready to move on to the next question, and it is the teacher who sets the pace for the exam.

The issue of not being able to go back and change a previous answer is one that the students mentioned:

‘You can’t go back to questions – you can’t stop and think’

‘In an oral when it’s done you should get a chance to go over it again. You should be able to write it down and then if you are stuck on it go through the rest of the exam and then come back to that question.’

‘You can take your time, but you can’t go back. There’s no second chance.’

In written exams students have to organise their own time and can go back to questions they miss out or think they might have got wrong. In an oral exam they cannot ‘leave a blank’ as they have little choice but to reply to the teacher with some sort of answer, and once they have said it they are more committed to it that they would be in a written exam where they could go back and change it.

The way in which time is managed in an oral exam may prove beneficial to low-achieving candidates such as those taking the CoA. By eliminating the need for them to manage their own time, it allows them to allocate more of their attentional resources to higher order skills necessary to answer the questions. We can describe the task in any exam as consisting of forming a strategy to answer questions, managing time, meeting the cognitive demands of the task, and an affective dimension. In the case of an oral exam the teacher removes the need for the students to form a strategy and apply it, and the need for time-management. This leaves the student freer to cope with affective aspects and to meet the cognitive demands of the task. The affective nature of the task in many exams includes the stress of writing and spelling, especially for low-achieving students. The oral exam removes this stress but it adds others. Although some aspects of the oral free up attentional processing for the demands of the task, others such as stress may have a negative affect on cognitive processing. The student-teacher relationship becomes more important in an oral exam, as does the confidence

of the student.

Confidence

There appear to be large individual differences in confidence amongst students in an oral exam. The more confident students will be better able to show the examiner what they know in an oral exam than the less confident students, whereas we would not expect to see such large discrepancies in a written exam. The questionnaires and interviews with teachers indicated that boys seemed more confident than girls in the oral exam.

‘Some were very confident, others very shy.’

‘Boys are more confident in oral than girls.’

‘Males are more confident.’

‘Girls were shy.’

‘Boys were more confident.’

These comments were backed up by evidence from the mean marks of girls and boys on the oral and written exams. Girls did better than boys in the written exam, as is the case with almost all subjects at this age. However, boys did better on the oral exam, and an Analysis of Variance showed a significant interaction between gender and performance on oral and written components ($F_{(2,50)} = 5.06, p < 0.03$). This may be due to the girls’ lack of confidence during the oral exam.

Sixty-five percent of the teachers felt that many of their candidates had been nervous or shy. One of the teachers remarked:

‘The assessment gave the opportunity but my candidates certainly did not show what they knew.’

This highlights the issue of the opportunities this kind of assessment provides for students. If students are able to take advantage of the opportunities given in the questions then the oral assessment allows them to show what they know. However, if they are nervous they may be unable to recognise the cognitive opportunities in the questions, and will not be able to demonstrate their level of knowledge of Geography.

We are left with the impression that it may be the interaction with a supportive second person that is more important than simply the removal of writing demands.

In some contexts, particularly the assessment of English as a Foreign Language, students are assessed orally in pairs in the belief that this helps those students who are too shy to talk to the teacher or assessor alone. There are complex issues involved in deciding who should be paired with whom and how this affects the assessment, but it is a possible method that the Geography teachers could use in a practice run or mock oral. Another way to reduce the effects of stress in an oral exam is for the teacher to use encouraging and reassuring language. In a detailed analysis of transcripts of 12 teachers’ oral exams we found that 11 of them used reassuring language throughout the assessment of their students.

The Use of Language

Conversational Implicature

The characteristics of the discourse in an oral exam are very different from those of ordinary language. Grice's notion of conversational implicature is useful here. Grice described our use of language as following a Cooperative Principle consisting of four maxims: Quality, Quantity, Relevance and Manner (see Levinson, 1983). The maxim of quality is that what we say is what we believe to be true, we say nothing that we believe to be false or for which we lack evidence. According to the maxims of quantity and relevance what we say is no more or less informative than necessary, and is relevant to the conversation. The maxim of manner states that we are brief and orderly in our conversation, without being ambiguous or obscure. According to these maxims, our conversation is efficient, rational and cooperative. The important point is that we assume that the other person in the conversation is also following the principle, so we make inferences about what they say, and these are conversational implicatures. The implicature of an utterance is the meaning intended by the speaker that is over and above the literal meaning.

According to this principle when someone asks a question this has the implicature that they do not know the answer and that they want to know the answer. If not, they would be violating the maxim of *quality*, so they would not be cooperating. One of the difficulties with the kind of conversation that is occurring in the Geography oral exam is that one of the participants – the teacher – is not adhering to the Cooperative Principle. The teacher is asking questions to which he or she *does* know the answer.

Young and He (1998) analysed Language Proficiency Interviews (LPIs) and found that the maxim of *quality* was present to a greater degree than that of *quantity*, the maxim of *relevance* was hardly used at all and the maxim of *manner* was causing the conversation to become verbose rather than clear. The LPI involves a conversation between a teacher and a student in a foreign language that the student is learning. The issues here are quite different from those in a Geography interview where it is not the conversation itself that is being assessed but the information conveyed in the conversation. If the Geography student is trying to convey everything that they know about the topic to the teacher in a short space of time during the interview then the nature of the discourse will differ considerably from ordinary language. *Quality* is likely to be present in an oral interview as the student is trying to tell the teacher what they know and so will endeavour to say nothing false. However, another aspect of *quality* is that we say nothing for which we lack evidence, and this may be violated in an exam situation in which students are unsure of an answer. Some students may be more inclined to risk saying something for which they have little evidence in an exam than they would in ordinary language, in order to try to get the mark. However, there will be individual differences here, and other students may be less likely to take such risks in an exam situation than they would in ordinary language because the exam makes them nervous. The way that 'exam stress' affects cognitive processing and discourse in oral exams is an issue that merits further investigation.

The fact that there is a 'right' answer in the teacher's mind during the oral has a considerable effect on the way that both the teacher and student use language. If the teacher indicates that the student has not given the answer expected, the student may be more likely to say something they are not sure about. The teachers on the other hand are starting by reading questions aloud, but are then moving on to elaborate on these questions, and this is when they are doing most of their talking. In most cases the teachers do more talking than the students during the oral.

The way in which the maxim of *quantity* is adhered to in the oral exam is also different from ordinary language. Some students offer more information than necessary in an attempt to 'hit on' the answer that is in the teacher's mind. However, most do not give any extra information and give very staccato responses. If they do give any extra information they often get a negative response from the teacher who tells them 'we will get to that later'. For example:

Student: 'But the main one's in Africa.'

Teacher: 'That was going to be my next question. Which part of the world has the largest area of rain forest at present?'

Student: 'South America. Er Africa, sorry.'

In this case the student gave extra but incorrect information in answer to the first question. The teacher then said that was the next question and went on to ask it. The student started by answering this one correctly but then reverted to the incorrect answer he gave earlier. The students very rarely say that they do not know the answer to a question. This would be a possible response in ordinary language, but it is less likely to occur in the oral exam in which the students want to try to gain credit. Again this is linked to the issue of *quality* – they will try to say something for which they lack evidence, whereas ordinarily they may say 'I don't know'. The students who do not wish to attempt a response tend to stay silent rather than say they don't know the answer. Silence would not be a socially acceptable response to a question in ordinary language but it occurs in the oral exam and is again linked to the students being nervous.

Relevance appeared to be almost non-existent in Young and He's (1998) analysis of LPIs. However, in the context of the Geography oral, most of the utterances are relevant. The students are not talking about other issues, but are trying to answer the question. The teachers read the question, and then perhaps repeat or re-phrase it, with the goal of getting the students to answer the questions and complete the oral.

The maxim of *manner* is also affected by the exam situation. When the teachers are reading the printed questions they are not speaking in a natural manner as some of the question wordings are vague and unclear. When the teachers go on to prompt students they often become verbose and start using complicated vocabulary to try to communicate the same question in a different way. For example:

Student: 'They're between the two tropics, along the Equator.'

Teacher: 'Right. Anywhere else? Any other descriptor you can give me?'

Student: 'Er, no.'

The teachers are trying to elicit as much information as possible from the students in answer to the questions, so they often repeat or re-phrase questions, and this, as well as lacking in brevity, often results in ambiguous questions and disorderly speech. Dickson (1981) showed that children respond more to the social implications of language than adults do, but as they become older and more educated they are able to differentiate the literal meaning of a statement from its implications. The students are therefore likely to be more sensitive than the teachers to the non-real language nature of the oral exam.

Interrogations

The Geography CoA oral exam appears to have nine of the twelve characteristics of interrogations listed by Low (1991). These nine characteristics are: a disproportionately large number of questions; questions are repeated often identically; there are certain desired or correct responses; conversational request sequences are not adhered to i.e. there are no apologies for the questions; there are fewer than the expected number of politeness formulae; the questioner changes topic without consulting the person questioned; the person being questioned is unable to change topic or control the direction of the interaction; there are

punishments for failures to answer correctly (no marks); and answers are formally recorded in writing or on tape.

There will be individual differences in how people respond to feeling they are being interrogated. Low (1991) suggests that those questioned are likely to avoid telling the questioner what they want to know if they feel they are being interrogated. They may also give vague rather than precise answers and even ignore certain questions entirely or respond only to one part of a question.

Although both participants in this sort of conversation tend to 'wear a mask', this can be dropped if the student feels that the teacher is on their side. If the teachers succeed in reassuring students and keep in sight the common goal, then the oral will not become an interrogation. Furthermore, as Low (1991) suggests, the teacher can lower the student's workload by recycling important information and marking it as important, thus reducing the density of the questions.

Conclusions

The oral exam in the Geography CoA has some distinct advantages over traditional written tests. Students who are unable to read or write well are given an opportunity to show what they know by another means. Teachers are able to interact with their students during the assessment, prompting them to answer to the best of their ability, and removing some of the strategic demands of the exam. This allows students to attend more directly to the cognitive demands of the task. Also, some of the students who do not do well on written tests are able to show what they know and gain a sense of positive achievement in the oral exam. It may be over simple to see the problems weak pupils face as only being related to the demands of writing, if the support they get in an oral exam in fact arises from the face to face interaction it involves.

Cognitive and Affective Factors

The disadvantages of the oral exam are also clear. Some students are under more stress than they would be in a written exam because they are nervous of speaking to their teacher in a taped exam situation. This can have an effect on their cognitive processing, causing them to devote attentional resources to coping with the stress. Any resulting incorrect processing of the questions or information in the resources can cause students to under-perform. Nerves seem to have affected the girls more than the boys in this study, but overall there will be individual differences in the level of stress experienced in oral exams. The issue of stress may be more of a problem for students taking the CoA than it would for other students if oral exams were to be used more widely. Teachers felt that students taking the CoA, who are low-achievers, are often very quiet in class, and lacking in confidence in comparison with other students. This again may be more true for the girls than the boys. If oral exams were to be used at different levels we might find that the affective dimension of the task influences the cognitive demands in different ways.

The Model for Assessment

The method of asking questions and then prompting to help the student or to ask for more information results in decreased reliability of the assessment. The teachers were not able to be consistent in the prompts that they gave to different students, as different students answered the questions in different ways and some needed more support than others. The issue here is whether we are intending to assess these students according to the hurdles they overcome in the task or according to the level of support they need to complete the task. Problems are inevitable if we fail to separate these two models.

Linguistic Issues

A further problem is associated with language use in the orals. Because there is a right answer in the teacher's mind that the student has to arrive at, the language differs from ordinary conversation, and the situation takes on aspects of an interrogation which can result in the student becoming disaffected. The teacher must ensure that the student feels they are on the same side, with a common goal.

Computer Based Assessment

One way of overcoming the problems of oral assessment without losing the benefits may be the use of a computer to present the questions. Students could answer questions on a computer, and the computer could prompt students according to their needs. In this way there can be rules for prompting, but questions can also be adapted and tailored for the particular student. This could remove the problems of confidence some students have in an oral test, and could also avoid the problem of assessing students in a linguistically unnatural conversation. If students could speak their answers into voice-recognition software, and the computer could speak the questions then the need for reading, writing and spelling would also be avoided. This may not be a natural situation for some people, but students are increasingly becoming used to interacting with computers in their learning, so it is plausible that this could be extended to assessment.

For students taking the CoA, the oral exam provides an opportunity for positive achievement which many of them may not find in traditional written exams. It also allows us to gain a clearer idea of students' understanding of a subject, although there are some students who are under too much stress to perform to their best ability. Oral exams have the potential for wider use, provided we develop a clear framework for how teachers should guide the students by prompting, and how this should be assessed. Students in the CoA do benefit from the teacher's prompts during the oral, and leave the exam with a sense of positive achievement. This prompting should be reflected in the marking criteria with a scheme of credit based on the amount of support needed.

There are, as mentioned earlier, serious problems in achieving adequate reliability and control of such a distributed process of assessment, difficulties that no examination board can afford to ignore. However, if these can be overcome, there is clearly potential for wider use of oral examining to help students show positively what they have learned and can do. Further investigation should concentrate on systematising the process without losing its potential.

References

Brown, G. & Yule, G. (1983) *Discourse Analysis*. Cambridge University Press.

Davies, B. & Corson, D. (Eds) (1997) *Oral Discourse and Education*. Encyclopaedia of Language and Education, Vol. 3. Kluwer.

Dickson, W.P. (Ed) (1981) *Children's Oral Communication Skills*. Academic Press.

Hartog & Rhodes (1936) *Marks of Examiners*. London.

Levinson, S.C. (1983) *Pragmatics*. Cambridge University Press.

Low, G. Talking to Questionnaires: Pragmatic Models in Questionnaire Design. In Adams, P., Heaton, B., & Howarth, P. (Eds) (1991) *Socio-cultural Issues in English for Academic*

Purposes. Review of English Language Teaching, Vol.1, No.2. Macmillan.
Pollitt, A. & Ahmed A., (1999) *A New Model of the Question Answering Process*. Paper presented at the International Association for Educational Assessment Conference, Slovenia, May 1999.

Utle, D., Mitchell, R. & Phillips, J.A. (1983) *Hear/say: a review of oral/aural graded tests*. Schools Council Examinations Bulletin, 44.

Young, R. & He, A.W. (Eds) (1998) *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency*. Studies in Bilingualism 14. John Benjamins.

“The question gets in the way: inferring competence from examination performance; some closing remarks” by Jackie Greatorex

University of Cambridge Local Examinations Syndicate

Closing remarks presented at the BPS London Conference, December 1999.

To consider what we have learnt about the complex structure of communication involved in the examination process and how examinations can be made fairer I will use the two parts of the title of the symposium;

- Does the question get in the way?
- Inferring competence from examination performance.

Throughout the symposium three other themes have emerged: -

- the quality of communication between participants in the examination process which affects the validity and fairness of the assessment;
- the cognitive processes and demands involved in different stages of the examination process;
- in some cases participants in the examination process fulfil more than one role in the triangle of communication between the setter, candidate and judge is a useful conceptualisation.

Does the question get in the way?

Task design has been illustrated to be an expert process and experts use 14 criteria to judge whether tasks are well designed. These criteria can be used to improve task design and thereby the communication between the setter and the candidate. One of these categories is at the heart of the questions raised in this symposium - ‘accessibility to students’ i.e. does the question fulfil its function of instructing the candidate about what they need to do. To determine the accessibility of the question/task to students it may be necessary to use both task setters, and students’ judgements and performance.

There are many factors which affect the quality of the communication between the setter and the candidate through the question. Paper 2 has focused on how reading and comprehending the question can be affected by factors such as time, stress, and specific features in the language. As has already been stated these factors can be a threat to the validity of the examination.

The research evidence suggests that different types of examinations (written and oral) and/or different types of questions place different cognitive demands on students. They each have their advantages and disadvantages. In both cases the quality of assessment depends on the communication between the setter and the candidate. In the case of oral examinations this is the communication between the teacher and candidate. Oral examinations are perhaps more flexible than written examinations since teachers can use prompting, but we also need tighter guidance on the use of prompting. There were a number of different types of prompts which were used and teachers found it difficult to be consistent. Throughout the oral examination the teachers’ role is to ask questions: so they also act as a setter adapting the task for the purposes of the candidate. Eventually the oral assessment may be a measure of how much help the assessor has given the candidate for them to get it right.

Inferring competence from examination performance.

When examiners decide on grade boundaries for each examination they are inferring competence from examination performance from current and archive scripts and the setters are communicating with the judges through the archive scripts. Additionally they are deciding

on the appropriate standard for each grade. Different groups of examiners may use different cognitive approaches for making these decisions (a categorisation task and a similarity judgement). There may be some similarities in the processes and further research may indicate which types of cognitive processes if any are more reliable. This in turn would facilitate good practice. It is important that examiners can infer competence and make fair judgements about standards for the examination process to be valid and reliable.

In summing up this symposium and beginning a general discussion I will draw upon the content of the papers and wider issues relating to examinations and psychology.

Improving education through evidence based assessment practice

The papers are an illustration of the usefulness of psychological research in improving education through assessment or in developing assessment for learning. Psychological research undertaken by examination boards and other research groups provides evidence for the practice of assessment. Educational assessment involves psychological measurement, but unfortunately there appears to be little psychological research in this area. One of the barriers to scientific research and its application in this area is that assessment in the UK is dominated by government policy, for example, the creation of the GCSE and the development of the National Curriculum. For an indepth discussion of how education has been increasingly politicised in recent years please see Lawton (1984). Educational assessment and examining in particular, is an area which would benefit from more psychological research.

At a recent British Educational Research Association Meeting of the Special Interest Group for Assessment it was suggested that some people in higher education (HE) were assessment illiterate (James, 1999). However there is a lot of research and development in HE about assessment, e.g. the reformation of the external examiner system and the 4th Northumbria Assessment Conference which took place in September at University of Northumbria, Newcastle upon Tyne. It is interesting to note that the work in these areas is often carried out by educational developers rather than psychologists. There may be room to improve assessment in higher education through psychological research and development, using what psychologists know about assessment as a basis for practice in HE. This suggestion has already been made by Newstead (1994).

This symposium has also illustrated the importance of cognitive research in the area of educational assessment. There are a variety of cognitive processes involved in examining, from the expert process employed by setters and judges to the cognitive processes used by the candidates and the cognitive demands of different questions and types of assessment. This is an area which would benefit from future research e.g.

- Examiners use archive scripts, statistics and other indicators to make their decisions. How do they use the information? Which information do they give priority to? Do they suffer from information overload?
- What are the different cognitive processes used by candidates to answer examination questions?
- What are the different cognitive demands of different types of assessment and tasks from oral to computer testing?
- What are the effects of new types of assessment on learning?

Standards

The buzzword in educational assessment and government education policy at the moment is 'standards'. In the case of national examinations like GCSEs and A levels examination boards are held publicly accountable for the maintenance of standards from year to year. There

appear to be increasing demands to monitor standards over time. Patrick (1996) gives four reasons for monitoring standards even when it is difficult to do so:-

1. Comparisons can be made between standards over short periods of time, say, four or five years, as during short periods of time the context does not change significantly. On the other hand it is not meaningful to compare standards over a much longer period of time because the context of tests will have changed significantly. Also these longer periods of time may include a discontinuity in the system like the introduction of the National Curriculum;
2. The monitoring may bring to light potential problem areas which could be investigated;
3. There may be subject or group trends and differences, for example, the changes in the achievement of girls and boys in recent years;
4. To highlight any changes in the curriculum.

She concludes that it is almost impossible to answer the question 'are standards rising or falling?' and her solution is to change the research question to the kinds of investigations listed above.

There may be room for more research in the area of maintaining standards and the changes in curriculum and patterns of achievement. For example, examination boards do not monitor any trends in achievement of young people from different ethnic groups or with special needs and this research may be worth undertaking.

The demand for the maintenance of standards implies that there are absolute standards which can be measured. This is not necessarily true. As an example level descriptors were developed as National Curriculum assessment criteria to ensure that the achievement of all students is measured against the same standards. Kimbell (1997) argued that useful level descriptors could not be developed because words did not define a standard; teachers interpreted the statements in terms of their experience of the students that they taught. He used the example of when the Design and Technology Working Group had defined level 10 (good performance expected of a sixteen year old) for the National Curriculum assessment criteria in 1988. When *primary* school teachers saw the descriptor they said "that's what my children do!".

It has been suggested that a standard is the aggregate of the 'demands' of the tasks which are being assessed. These demands can vary from, cognitive demands like problem solving or understanding the question to more emotional demands like stress management. The interaction of the demands of the task is expressed in the candidates' performance. Ideally candidates will be assessed on their performance on valid tasks, for example is it valid to assess candidates' ability to write an essay when they are unlikely to write an essay after they have left education? The type of tasks which should be used to assess students depends upon the purposes of the assessment. The problem with developing educational tasks for assessment purposes is that there are many purposes and users of a test. For example, GCSEs are for employers to help decide whether a candidate is worth employing, teachers to assess whether the candidates can go on to further study and now the government is proposing that they should be used as diagnostic as well as summative forms of assessment to help students to identify their shortcomings and improve their learning.

There are increasing requests from teaching and research professionals to consider the psychological effects of the heavy assessment demands that are placed on students at an increasingly younger age (Threlfall, 1999). Indeed there is a suspicion that in some circumstances assessment may hinder rather than help learning, especially in the case of very young children. This is another important area for future psychological research.

Conclusion

Educational assessment and the public examination system is a valid area of psychological investigation where more research is to be welcomed.

References

James, D. (1999). *Assessment in Higher Education*, a paper presented at the Special Interest Group in Assessment, 27 May, University of Leeds.

Kimbell, R. (1997). *Assessing Technology, International trends in curriculum and assessment*. Open University Press: Buckingham.

Lawton, D. (1984), *The Tightening Grip: the growth of central control of the school curriculum*, Bedford Way, Paper No 21 ULIE.

Newstead, S. E. (1996). The Psychology of Student Assessment. *The Psychologist*, 9 543 - 547.

Patrick H. (1996) *Comparing Public Examination Standards over Time*, a paper presented at the British Educational Research Association Conference, University of Lancaster.

Threlfall, S. (1999) *The Practitioner's Tale*, a paper presented at the Special Interest Group in Assessment, 27 May, University of Leeds.