Department for Education and Skills (2005). *14–19 Education and Skills*. London: The Stationery Office.

Department of Education and Science (2005). Irish educational information available online at http://www.education.ie/home/

Eurostat (2005). Detailed statistics on the EU and candidate countries available online at http://epp.eurostat.cec.eu.int

Fuller, A. & Unwin, L. (2005). Opinion: Vocational guidance. *Guardian Unlimited*, Tuesday 29th March.

Harris, M. (2003). *Modern apprenticeships: An assessment of the Government's flagship training programme*. Policy paper. London: Institute of Directors.

Hodgson, A. & Spours, K. (2005). Divided we fail. *The Guardian*, Tuesday 1st March.

Johnson, M. (2005). *Vocational review: Issues of validity, reliability and accessibility*. Internal report for Cambridge Assessment.

Organisation for Economic Co-operation and Development (2005). *Education at a glance*. Available online at www.oecd.org/edu/eag2005

Regional Education Indicators Project – PRIE (2002). *Regional report: Educational panorama of the Americas*. Santiago, Chile: UNESCO Regional Office for Education in Latin America and the Caribbean.

Thomson, P., Mathers, R. & Quick, R. (1996). *The grade debate*. Adelaide, Australia: National Centre for Vocational Education Research.

Tomlinson, M. (2004). *14–19 Curriculum and qualifications reform: Final report of the working group on 14–19 reform*. London: DfES Publications.

UNESCO (1997). International Standard Classification of Education (ISCED). Available online at www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

UNESCO (2003). Institute for Statistics. Data available online at www.uis.unesco.org

Universities and Colleges Admissions Service (2005). *International qualifications 05: for entry into higher education*. Cheltenham: UCAS.

Unwin, L., Fuller, A., Turbin, J. and Young, M. (2004). *What determines the impact of vocational qualifications?* Department for Education and Skills Research, Report No 522.

Williams, M. & Bateman, A. (2003). *Graded assessment in vocational education and training: An analysis of national practice, drivers and areas for policy development*. Leabrook, Australia: National Centre for Vocational Education Research.

PSYCHOLOGY OF ASSESSMENT

A cognitive psychological exploration of the GCSE marking process

Dr Irenka Suto and Dr Jackie Greatorex Research Division

Background

GCSEs play a crucial role in secondary education throughout England and Wales, and the process of marking them, which entails extensive human judgement, is a key determinant in the futures of many sixteen-year-olds. While marking practices in other kinds of examinations have received some serious consideration among researchers (for example, Cumming, 1990; Vaughan, 1992; Milanovic *et al*., 1996; Laming, 1990, 2004; Webster *et al*., 2000; Yorke *et al*., 2000), the judgements made during GCSE examination marking remain surprisingly little explored. The aims of our study, therefore, were to investigate the cognitive strategies used when marking GCSEs and to interpret them within the context of psychological theories of human judgement.

Within the broad field of psychology, there exist multiple models of judgement and decision-making, which have yet to be applied to GCSE examination marking. One potentially useful theoretical approach is that of dual processing. Such models distinguish two qualitatively different but concurrently active systems of cognitive operations: *System 1* thought processes, which are quick and associative, and *System 2* thought processes, which are slow and rule-governed (Kahneman and Frederick, 2002; Stanovich and West, 2002).
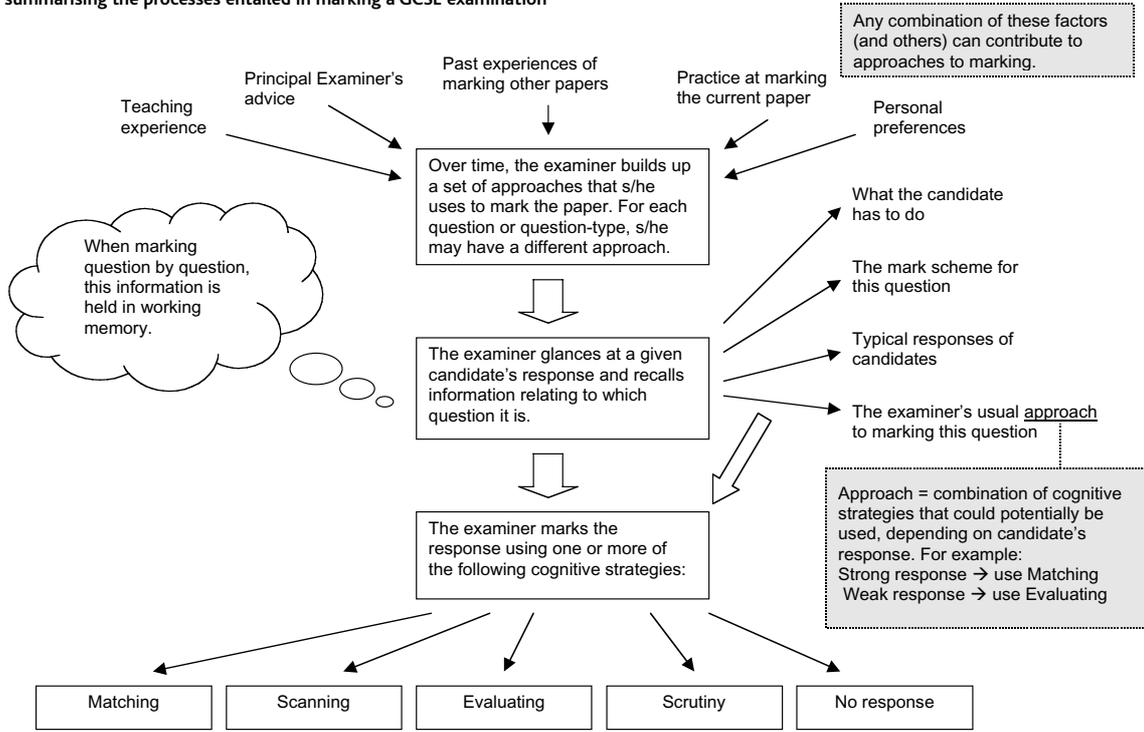
The 'intuitive' judgements of System 1 are described as automatic, effortless, skilled actions, comprising opaque thought processes, which occur in parallel and so rapidly that they can be difficult to elucidate (Kahneman and Frederick, 2002). System 2 judgements, in contrast, have been termed 'reflective', and the thought processes they comprise are characterised as slow, serial, controlled, and effortful rule applications,

of which the thinker is self-aware (*ibid*. 2002). According to Kahneman and Frederick (2002), as an individual acquires proficiency and skill at a particular activity, complex cognitive operations may migrate from System 2 to System 1. For example, chess masters can develop sufficient expertise to perceive the strength of a chess position instantly, as pattern-matching replaces effortful serial processing.

GCSE examination marking is a diverse activity, encompassing a wide range of subjects with a variety of question styles and mark schemes. It is likely, therefore, that at least some aspects of it will have parallels with some of the activities already scrutinised by judgement researchers in other contexts. There may be question types, or stages of marking, that involve System 1 processing; at times, simple and repetitive matching of a candidate's single-word response with the model answer given in the mark scheme may be all that is required. At other times, examiners might be engaged in System 2 processing; for example, when carefully applying the complex guidelines of a mark scheme to a candidate's uniquely worded essay. As examiners become more familiar with a particular examination paper and mark scheme, or more experienced at marking in general, some sophisticated thought processes may be transferred from System 2 to System 1, while others remain exclusive to System 2.

In the present investigation, we sought to identify and explore some of the many judgements made by GCSE examiners. To do this, we conducted a small-scale empirical study of examiners marking two contrasting subjects, in which we used the 'think aloud' method (Ericsson and Simon, 1993; Leighton, 2004; Van Someren *et al*., 1994) to obtain verbal protocol data for qualitative analysis.

Figure 1: Model summarising the processes entailed in marking a GCSE examination

Methods

Two GCSE examinations (administered by OCR) were considered: an intermediate tier Mathematics paper, which used a 'points-based' marking scheme, and a foundation tier Business Studies paper, which used a 'levels-based' scheme. For both examinations, candidates' scripts comprised individual booklets containing subdivided questions with answer spaces beneath each question part.

For each subject, a group of six experienced examiners (one Principal Examiner and five Assistant Examiners) marked four identical script samples each. The first three of these samples were marked silently (for details, see Suto and Greatorex, *in press*). They were used to familiarise the examiners with the papers and coordinate their marking. Whilst marking the fourth sample (comprising five scripts), the examiners were asked to 'think aloud' concurrently, having been instructed: '…*Say out loud everything that you would normally say to yourself silently whilst you are marking…*' Using a semi-structured interview schedule, the examiners were later questioned about their marking experiences retrospectively.

Results

An extensive qualitative analysis, and interpretation, of the verbal protocol data enabled us to propose a tentative model of marking, which includes five distinct cognitive marking strategies: *matching, scanning, evaluating, scrutinising,* and *no response*. An overview of the model is presented in Figure 1, and the five strategies are presented in detail in Figures 2 to 6. (There is a key to these figures on page 9.) These strategies were broadly validated not only in the retrospective interviews with the examiners who participated in the study, but also by other senior mathematics and business studies examiners.
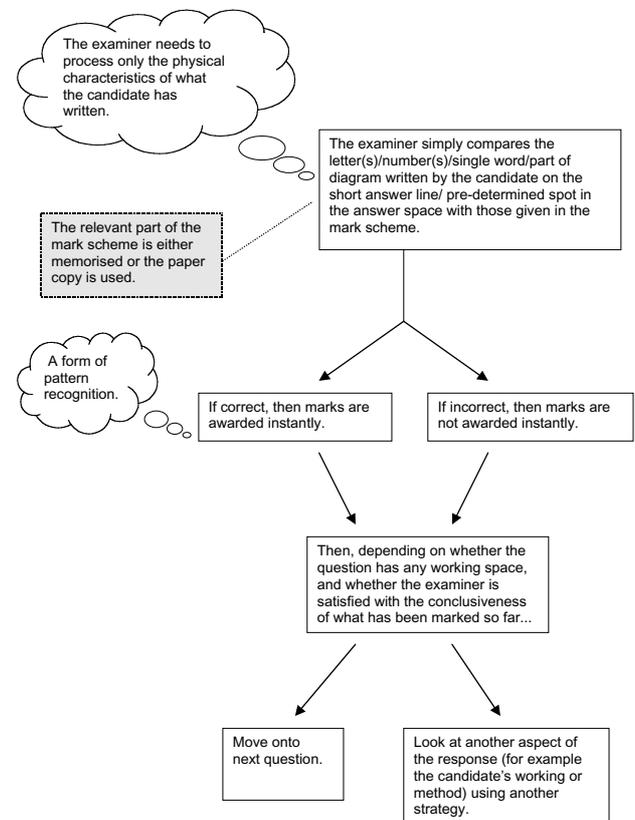
Figure 2: The 'Matching' strategy
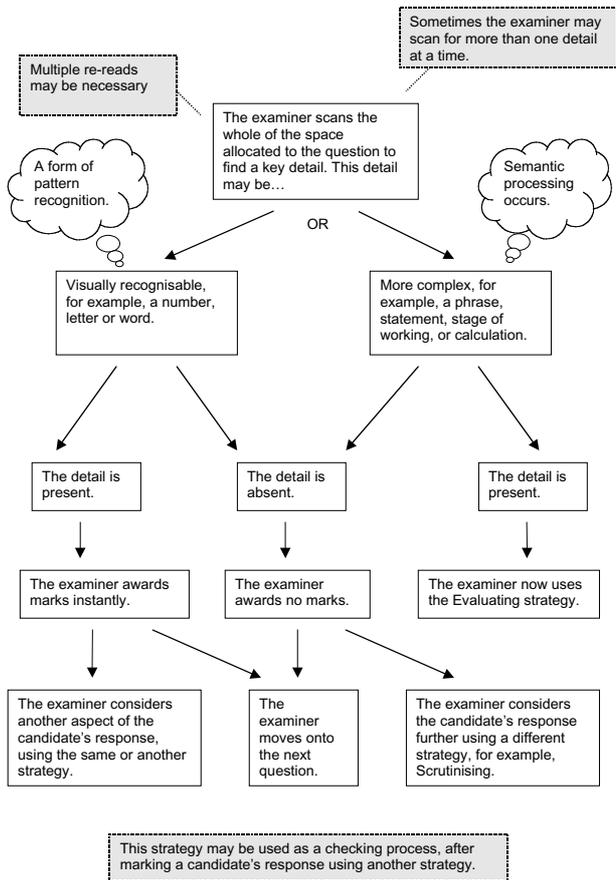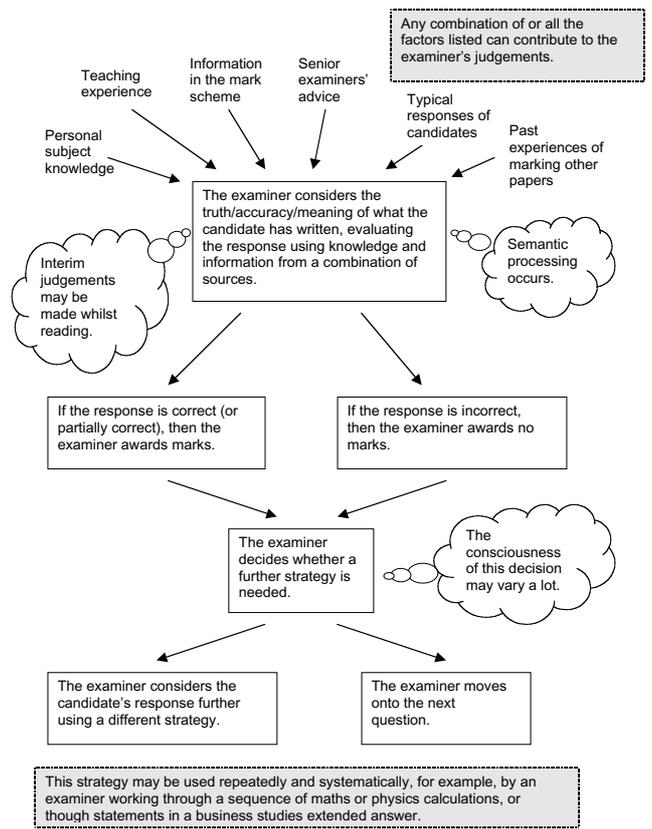
Figure 3: The 'Scanning' strategy

Multiple re-reads may be necessary

Sometimes the examiner may scan for more than one detail at a time.

The examiner scans the whole of the space allocated to the question to find a key detail. This detail may be…

A form of pattern recognition.

Semantic processing occurs.

OR

Visually recognisable, for example, a number, letter or word.

More complex, for example, a phrase, statement, stage of working, or calculation.

The detail is present.

The detail is absent.

The detail is present.

The examiner awards marks instantly.

The examiner awards no marks.

The examiner now uses the Evaluating strategy.

The examiner considers another aspect of the candidate's response, using the same or another strategy.

The examiner moves onto the next question.

The examiner considers the candidate's response further using a different strategy, for example, Scrutinising.

This strategy may be used as a checking process, after marking a candidate's response using another strategy.

Figure 4: The 'Evaluating' strategy

Teaching experience

Information in the mark scheme

Senior examiners' advice

Any combination of or all the factors listed can contribute to the examiner's judgements.

Personal subject knowledge

Typical responses of candidates

Past experiences of marking other papers

The examiner considers the truth/accuracy/meaning of what the candidate has written, evaluating the response using knowledge and information from a combination of sources.

Interim judgements may be made whilst reading.

Semantic processing occurs.

If the response is correct (or partially correct), then the examiner awards marks.

If the response is incorrect, then the examiner awards no marks.

The examiner decides whether a further strategy is needed.

The consciousness of this decision may vary a lot.

The examiner considers the candidate's response further using a different strategy.

The examiner moves onto the next question.

This strategy may be used repeatedly and systematically, for example, by an examiner working through a sequence of maths or physics calculations, or though statements in a business studies extended answer.

Figure 5: The 'Scrutinising' strategy

When a response is unexpected and/or wrong and/or not the same as any of those given in the mark scheme, the examiner may need to establish precisely where the problem lies, or whether the response is actually a correct and valid alternative. The examiner's overall aim is to reconstruct the candidate's line of reasoning or establish what the candidate has attempted to do. This entails:

This strategy is not used alone. It follows on from, or is used together with, other strategies.

Multiple re-reads may be necessary.

The overall aim of attempting to reconstruct the candidate's line of reasoning or establish what the candidate has attempted to do.

AND/OR

Searching for particular kinds of errors or inconsistencies in the candidate's working and reasoning.

The examiner decides whether the response deserves any marks, and awards them accordingly.

The examiner uses a further strategy, for example, Scanning as a checking process.

Move onto next question/ question part.

Figure 6: The 'No response' strategy

No semantic processing is needed.

When the candidate appears to have written nothing at all, the examiner looks over the space allocated to the questions more thoroughly to confirm this.

The examiner discovers that the candidate has actually written something.

Nothing has been written, so the examiner awards no marks and moves on to the next question.

The examiner uses another strategy to decide whether to award any marks.
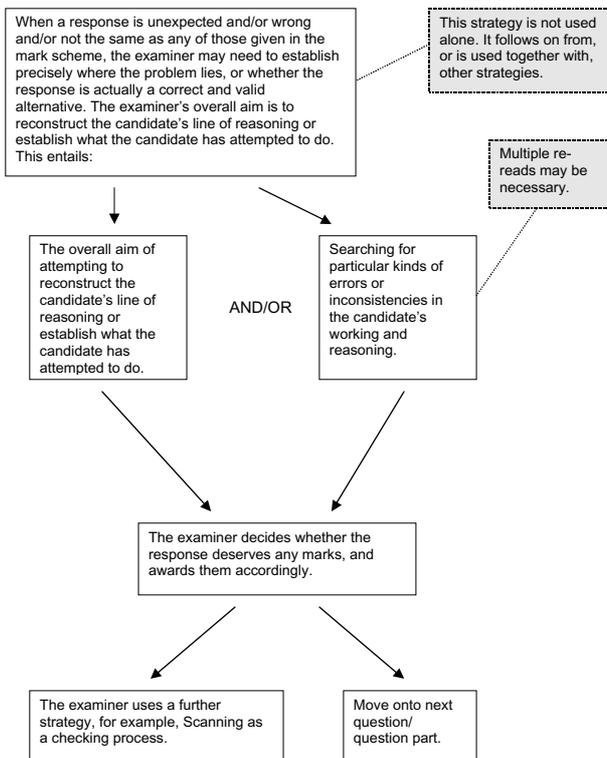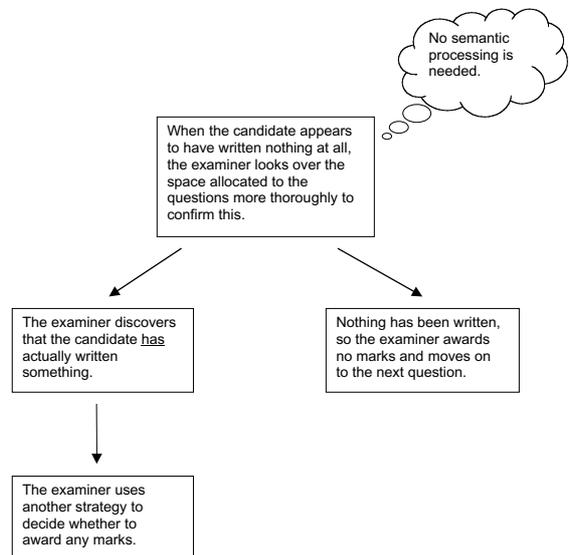
Key to Figures 1 to 6

These boxes describe what the examiner is doing.

These bubbles indicate how the strategy relates to psychological phenomena.

These boxes provide some additional notes about the strategies

Figures 7 and 8 contain transcript excerpts that can be taken to illustrate the five cognitive marking strategies. As these figures indicate, the marking strategies can be interpreted within dual-processing theories of judgement as comprising essentially System 1 judgement or System 2 judgement, or even both.

Figure 7: Examples of Mathematics examiners using each strategy

Verbal protocol excerpt	Strategy used	System 1('intuitive') or System 2 ('reflective') judgements?
Number seventeen: 61.49 instead of sixty-nine pounds seventy, so no marks there.	Matching	System 1
While I do that I'm sort of staring at the page, and I can see a four sixty-eight. Whoopee!	Scanning	System 1
We're looking for them using sine to find the angle.	Scanning	System 2
His algebra is fine.	Evaluating	Both Systems 1 and 2
Ugh, yuk. What a mess... Ah, looking at a whole load of percentages, all over the place. Er, and...that's a calculator paper. You wouldn't know it 'cause they're using non-calculator methods...and...that for division by a hundred. Can't...do it without doing a long division, poor dears. But it's all a load of... I'm trying to, erm, get close. It's all in trial and improvement... Erm, M0 for that one. Trying to, don't know how to divide things.	Scrutinising	Both Systems 1 and 2
And part E: no response so that gets nothing.	No response	System 1

Figure 8: Examples of Business Studies examiners using each strategy

Verbal protocol excerpt	Strategy used	System 1 ('intuitive') or System 2 ('reflective') judgements?
Four two five three is the answer. Four...no, no, nope, nope. No marks.	Matching	System 1
The answer on this one is 'Rea Aitkin, chairman', so as soon as I see 'Rea Aitkin, chairman', it's two marks.	Scanning	System 1
And looking for an action by Belgian chocolate manufacturers...	Scanning	System 2
The community, a judgement is made that the community should be considered, and a reason is that because they are consumers and obviously that would affect sales. Only a simple answer: one mark.	Evaluating	Both Systems 1 and 2
Now unusually, this candidate is suggesting, er, Miss Singh...as a decision-maker. I'm just checking the...this is the finance director. Erm, I'm accepting that, because a finance director can, sometimes, hold key decision-making influence. Er, I'm looking for a reason. 'He deals with the finances.' That's okay. I've accepted that for two marks.	Scrutinising	Both Systems 1 and 2
Blank: nothing.	No response	System 1

Discussion

The aims of this study of GCSE examiners were to identify the key marking strategies used, and to interpret them within the context of dual processing theories of judgement. There were several limitations, which included: the use of small samples of examiners; the exploration of just two GCSE examinations; four examiners not managing to mark all of their scripts; qualitative analysis inevitably involves some interpretation by the researchers and the potential of the process of 'thinking aloud' to interfere with the thought processes under investigation (for example, slowing them down). Together, these restrictions mean that our model of strategies is unlikely to be exhaustive.

Nevertheless, our study has some important implications. First, the complexity of some of the strategies identified confirms that GCSE examination marking can be a cognitively demanding process, often requiring considerable expertise. For some questions, the simpler strategies could, arguably, be used by many people, including those without much expertise and experience. However, those strategies that rely on subject knowledge, past marking and/or teaching experience, and on advice from the Principal Examiner, for example, *evaluating* and *scrutinising*, are often necessary when a candidate's response is long or unexpected.

Secondly, knowledge of the strategies identified in this study may prove useful to senior examiners. While several examiners have suggested that our named strategies provided a useful language with which to communicate with colleagues, others have suggested using the research in training courses for new examiners.

Thirdly, the study provides grounds upon which to hypothesise that some of the judgements entailed in marking may start off as slow and conscious System 2 thought processes, but migrate to System 1 as an examiner either acquires expertise or gains confidence. Examiners who were interviewed about the study supported this hypothesis, and several raised concerns about some examiners switching from using System 2 to using System 1 on particular questions before they were ready to do so. Several individuals felt that knowledge of the strategies would provide a means of 'self-checking' for all examiners, who could thereby remind themselves periodically of the need to evaluate and scrutinise some responses.

Finally, explicit knowledge of the strategies could prove useful when designing examination papers and mark schemes. For example, although it is impossible to predict every potential answer to a given question, listing as many valid responses as possible in the form of bullet points when the *matching* strategy is most likely to be used, or listing key information to scan for where a scanning strategy is viable, could help maximise the efficiency of the marking process.

Further reading

A full report of the study described here is soon to be published in the *British Educational Research Journal* as an article entitled 'What goes through an examiner's mind? Using verbal protocols to gain insights into the GCSE marking process'.

References

Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, **7**, 31–51.

Ericsson, K. & Simon, H. (1993). *Protocol Analysis: Verbal reports as data*. London: MIT Press.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Laming, D. (1990). The reliability of a certain university examination compared with the precision of absolute judgments. *Quarterly Journal of Experimental Psychology*, **42A**, 239–54.

Laming, D. (2004). *Human judgment: The eye of the beholder*. London: Thomson.

Leighton J. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, Winter, 6–15.

Milanovic M., Saville, N. & Shuhong, S. (1996). A study of the decision-making behaviour of composition markers. In M. Milanovic & N. Saville (Eds.). *Studies in Language Testing 3: Performance testing, cognition and assessment –*

Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem. Cambridge: Cambridge University Press/University of Cambridge Local Examinations Syndicate.

Stanovich, K. & West, R. (2002). Individual differences in reasoning. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Suto, W.M.I. & Greatorex, J. (*in press*). What goes through an examiner's mind? Using verbal protocols to gain insights into the GCSE marking process. *British Educational Research Journal*.

Van Someren, M., Barnard, Y. & Sandberg, J. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

Vaughan, C. (1992). Holistic assessment: what goes on in the rater's mind? In L. Hamp-Lyons (Ed.). *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex.

Webster, F., Pepper, D. & Jenkins, A. (2000). Assessing the undergraduate dissertation. *Assessment and Evaluation in Higher Education*, **25**, 71–80.

Yorke, M., Bridges, P. & Woolf, H. (2000). Mark distributions and marking practices in UK higher education. *Active Learning in Higher Education*, **1**, 7–27.

PSYCHOLOGY OF ASSESSMENT

Examiners' annotations: Practice and purpose

Victoria Crisp and Martin Johnson Research Division

Introduction

'*When you come to any passages that seem to you useful, make a firm mark against them, which may serve as lime in your memory, less otherwise they might fly away.*'
Advice from St Augustine in Petrarch: *Secretum Meum* 1358

The processes of reading and writing are recognised to be inextricably intertwined. Writing helps to support cognitive demands made upon the reader whilst processing a text (e.g. O'Hara, 1996; Benson, 2001). Anderson and Armbruster (1982) suggest that annotating activities are concurrent with the actual reading processes, influence the way that reading occurs, and the way that meaning is processed. Examiners annotate scripts whilst marking (e.g. underlining, circling, using abbreviations or making comments) and this may reflect the cognitive support for comprehension building that annotations can provide.

Within the accountability agenda that pervades education there is an emphasis on clear communication channels between examiners of different seniority to facilitate effective monitoring. Annotations might have an important communicative role in this quality control process by offering others up and down the chain an insight into the rationale behind the annotating examiners' decisions. Previous re-marking investigations have suggested that annotations do have a communicative function, potentially influencing how subsequent viewers perceive the quality of a script (Murphy, 1979; Wilmut, 1984; Newton, 1996). Laming (2004) suggests that this is because there are places where the mark scheme leaves the examiner uncertain, and that judgements in such cases are influenced by extraneous information, for example, the previous annotations of other judges.

In addition to evidence that annotations act as a communicative device, there is also evidence that annotating might have a positive influence on markers' perceptions and affect their feelings of efficacy. Most markers felt that annotating improved their marking, helping them to apply performance criteria and reducing the subjectivity of judgements (Bramley and Pollitt, 1996). In pilot work on online assessment teachers, examiners and moderators have expressed dissatisfaction where facilities for annotation were limiting (Greatorex, 2004; Raikes *et al.*, 2004). Markers report that using annotations provides an efficient means to confirm or reconsider standards both within and across candidates as well as acting as a reassurance during the judgemental process (Shaw, 2005).

Rationale

The literature available provides some information about the purposes and effects of annotations. However, there is a relative sparsity of published research about annotation in examination marking in terms of the following:

● consistency of use of codes

● examiners' reasons for using annotations

● the role that annotations might be playing in decision making processes

● the effects, or perceived effects, of using annotations whilst conducting first marking.

This research investigates some of these issues and develops a more comprehensive picture of annotation practices.