

The Locals began in the 1850s partly because they were seen to be a fair way to identify and reward ability. Twenty years later the boards were working out their responses to accusations of not being fair. It is how far the present examination system is seen as being fair to individuals which will make the difference between it being perceived as a liberalising or a reactionary one. Independent examining boards are well-placed to respond to this challenge and they continue to play an important part in maintaining the credibility of that system.

References

Bentham, J. (written and printed 1827, published 1830). *Constitutional code: Vol. 1*. London: R.Heward.

English Journal of Education (1857). XI, 226.

Foden, F. (1989). *The Examiner, James Booth and the Origins of Common Examinations*. Leeds: School of Continuing Education, University of Leeds.

Headmistress's Reports to Governors (1871). II, 256–256.

Herbert, A. (ed.) (1889). *The sacrifice of education to examination*. London: Williams & Norgate.

Latham, H. (1877). *On the action of examinations considered as a means of selection*. Cambridge: Deighton, Bell.

Mill, J.S. (1859). *On Liberty*. London: John W. Parker & Son.

Montgomery, R.J. (1965). *Examinations: An account of their evolution as administrative devices in England*. London: Longmans.

Roach, J. (1971). *Public Examinations in England: 1850–1900*. Cambridge: Cambridge University Press.

Royal Commission on Secondary Education 1894. (1895). V., 268. London: HMSO.

Sandford, E.G. (ed.) (1906). *Memoirs of Archbishop Temple by 'Seven Friends'*. II, 541, 68–69; 549. London: Macmillan & Co. Ltd. Quoted in Roach (1971).

Sessional Proceedings of the National Association for the Promotion of Social Science, 1871–72. 29.

Stephen, B. (1927). *Emily Davies and Girton College*. 83–4. London: Constable. Quoted in Roach (1971).

The Transactions of the Social Science Association (1872). 271–9

Weise, L. (trans. L. Schmitz, 1877). *German Letters on English Education: Written during an Educational Tour in England in 1876*. London: Collins.

Woodard, N. (1848). *A Plea for the Middle Classes*. (pamphlet).

Woodard, N. (1851). *Public Schools for the Middle Classes*. (pamphlet).

PSYCHOLOGY OF ASSESSMENT

Investigating the judgemental marking process: an overview of our recent research

Dr Irenka Suto, Victoria Crisp and Dr Jackie Greatorex Research Division

Prior to Cambridge Assessment's recent interest in the area, the process of marking GCSE and A-level examination questions had received surprisingly little attention among psychologists and social scientists. Whilst there has been some research into marking processes in other contexts (e.g. English essay marking: Barritt, Stock and Clark, 1986, Pula and Huot, 1993; English as a second language: Cumming, 1990, Vaughan, 1991, Milanovic, Saville and Shuhong, 1996, Lumley, 2002) to our knowledge, only Sanderson (2001) has explored the process in depth, producing a socio-cognitive model of A-level essay marking. To address this dearth in knowledge, members of the Core Research team have conducted several linked projects, considering the marking process from different angles. Key aims have been to provide insights into how examiner training and marking accuracy could be improved, as well as reasoned justifications for how item types might be assigned to different examiner groups in the future.

As with any major research question, the issue of how examiners mark items needs to be explored from many different angles to gain as full and cohesive an answer as possible. In biological research, for example, the nature and effects of an illness are explored at multiple levels: molecular, intra-cellular, cellular, physiological, whole organismal, and even epidemiological and demographic. Similarly, some physics researchers conduct fine-grained analyses of minute particles, while others monitor massive structures in space, both in their attempts to establish how the universe began. Linking together these jigsaw pieces in order to see the bigger picture and gain a real overview of a process or phenomenon can be a difficult but necessary challenge. As with biology and physics, this is

an important task for researchers in educational assessment.

To recognise the different approaches to research and analysis that the marking process engenders, it is worth considering the very broad research field in which it primarily lies – that of human judgement and decision-making. There exist a number of well-established approaches to investigation, adopted by researchers working within diverse paradigms, and as with the natural sciences, questions are explored on a number of levels. For example, a key approach has been to ask *what information* people attend to and utilise when making decisions. On perhaps the most 'fine-grained' level of research, cognitive psychologists have identified and scrutinised shifts in visual attention among small pieces of information, such as letters, numbers and words on a page of writing. At another level, other psychological researchers have focused on cognitive heuristics and biases in information processing. At yet another level, the influences and roles of behavioural and social information have been explored by researchers interested in such dynamics, and at yet another level still, the effects of physical information in the environment have been studied. Studies at all of these levels have provided important contributions to our understanding in the research field, although there is the potential for them to be integrated much more.

Another popular approach to understanding judgement and decision-making has been to explore the sequences of mental operations by which people arrive at their choices. This approach has proven particularly popular in clinical and legal settings, and again, it has been adopted at a number of levels. In the Core Research team's work on the marking process, we have combined this approach with the one outlined above:

our projects explore both the information that people attend to when marking items and the sequences of mental operations involved.

At a relatively fine-grained level, in our first project, entitled *Markers' Minds*, we identified the cognitive marking strategies entailed in marking GCSE maths and business studies questions (Suto and Greatorex, *in press*, a). This was done using the think aloud method with experienced examiners (Greatorex and Suto, *in submission*). Working within the 'dual-processing' paradigm (Kahneman and Frederick, 2002), we interpreted two of our five strategies (*matching* and *no response*) as utilising simple 'System 1' or 'intuitive' judgemental processes, two strategies (*scrutinising* and *evaluating*) as utilising more complex 'System 2' or 'reflective' processes, and one strategy (*scanning*) as engaging System 1 and/or System 2. An analysis of strategy usage (Greatorex and Suto, 2005; Suto and Greatorex, *in press*, b) revealed that although there are some differences among individual examiners, the most prominent differences occur between subjects and among questions.

A second closely-linked project (Greatorex and Suto, 2006) entailed re-analysing the verbal protocols of 'expert' and 'subject' markers who had marked GCSE maths and A-level physics papers on screen. The data provided further evidence for the five cognitive strategies, indicating that they are used in other marking contexts.

As part of a subsequent project, entitled *Marking Expertise 1*, we explored the relationship between cognitive marking strategy complexity and marking accuracy (Suto and Nádas, 2007a; Suto and Nádas, *in press*). A new theoretical framework was constructed, conceptualising marking accuracy for a particular question as being determined by both (i) a marker's expertise, and (ii) the demands of the marking task. It was proposed that these two factors are in turn affected by a number of different variables, including the complexity of the marking strategies needed to mark the question, and utilisation knowledge of the marking strategies (i.e. knowing which strategies to apply when).

The question-by-question marking of GCSE maths and physics by

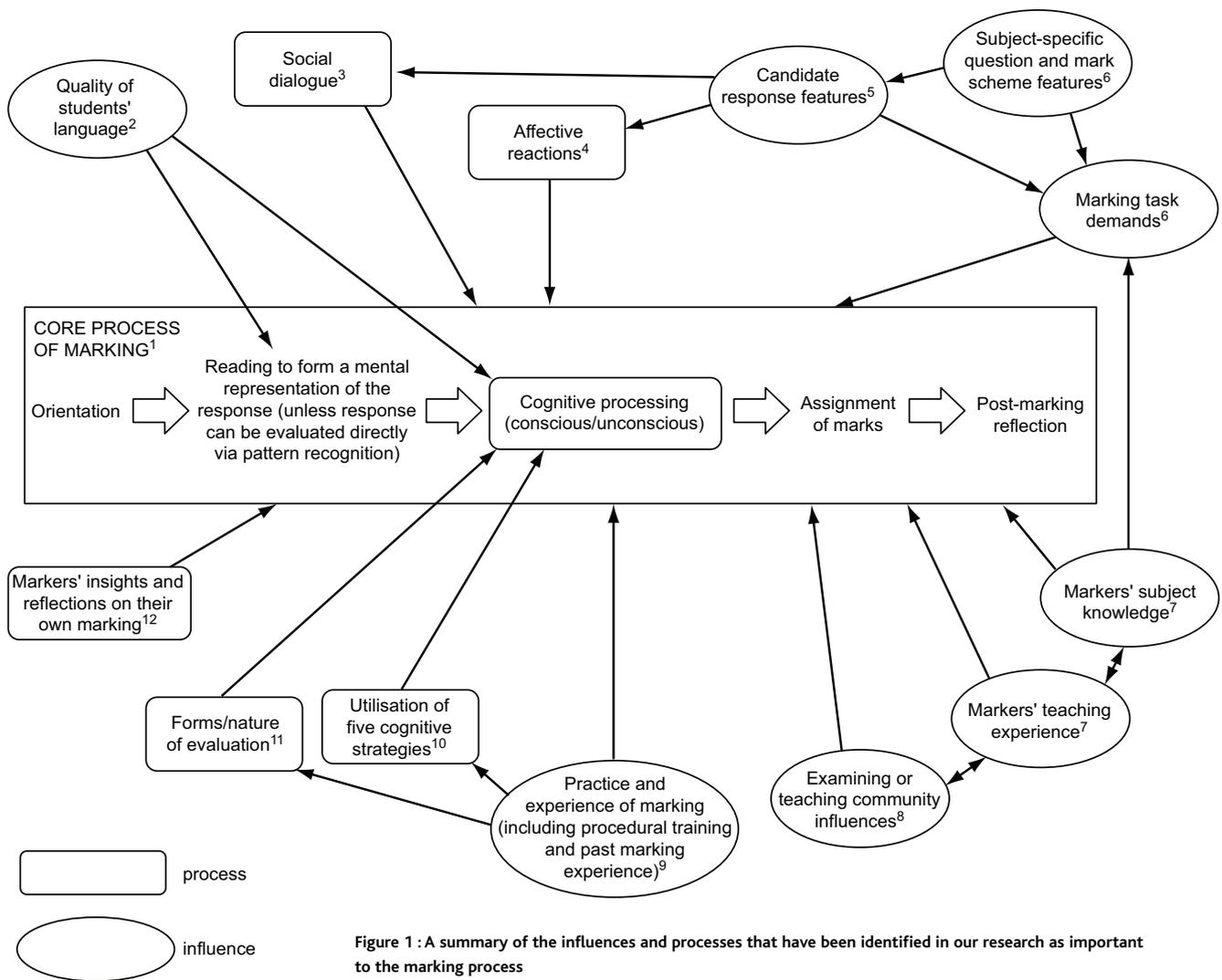


Figure 1 : A summary of the influences and processes that have been identified in our research as important to the marking process

1. Crisp (2007c; *in submission*), Greatorex and Suto (2006), Suto and Greatorex (2006)
2. Crisp (2007a; 2007b; 2007c)
3. Crisp (*in press*)
4. Crisp (2007b; *in press*)
5. Crisp (2007a; 2007b; *in press*; *in submission*)
6. Suto and Nádas (2007a; 2007b)
7. Greatorex, Nádas, Suto and Bell (2007), Suto and Nádas (2007a, *in press*)
8. Crisp (*in submission*), Greatorex, Nádas, Suto and Bell (2007), Suto and Nádas (2007a, *in press*)
9. Greatorex and Bell (*in press*), Greatorex, Nádas, Suto and Bell (2007), Suto and Nádas (2007a, *in press*)
10. Greatorex (2006; 2007), Greatorex and Suto (2006), Suto and Greatorex (*in press*, a, b)
11. Crisp (2007b; 2007c; *in press*; *in submission*)
12. Nádas and Suto (2007), Suto and Nádas (2007a)

'expert' and 'graduate' markers was explored, and it was found that the apparent marking strategy complexity that a question entails is indeed associated with the accuracy with which it is marked. (Instead of using the think aloud method to demonstrate which marking strategies were used, researchers rated the marking strategy complexity of each question a priori.) The finding was replicated in a study of A-level biology marking (*Internalising Mark Schemes*, reported as part of Greateorex, Nádás, Suto and Bell, 2007). Furthermore, one finding was that apparent marking strategy complexity was a useful indicator of how much the standardisation meeting improves marking accuracy; this was the case for two of the three subjects investigated (Greateorex *et al.*, 2007). In *Marking Expertise 1*, apparent marking strategy usage was also found to be associated with various subject-specific question features, which are in turn associated with accuracy (Suto and Nádás, 2007b).

These projects have been generally well received, and researchers outside Cambridge Assessment who attended conference presentations of the *Markers' Minds* research have been interested to know how the cognitive strategies relate to other more socio-cognitive perspectives. This question has begun to be addressed in another project: *Holistic versus Structured marking* (Crisp, 2007a; 2007b; *in press; in submission*). The primary aim of this research was to compare the process of marking short/medium answer questions with that of marking essays. This was achieved in the context of A-level geography, again using the think aloud method to collect data from examiners. This time, however, the analysis was broader, covering a number of different levels. Several well-established theoretical perspectives were brought into the analysis: constructivist theories of reading comprehension, discourse communities, and communities of practice.

A number of types of examiner behaviours and reactions were identified which were compared between question types within the qualification (Crisp, 2007a; Crisp, *in press*). The framework was also used to explore individual differences among examiners, a considerable number of which were revealed. Possible associations between marker behaviours and lower marker agreement were investigated leading to tentative implications (Crisp, 2007a; Crisp, *in press*). The appropriateness of the features that examiners attended to was also analysed (Crisp, 2007b). A broad socio-cognitive model bringing together the behaviours and reactions observed was proposed to represent the phases (and loops) involved in the process of marking responses. Links between the proposed model and existing psychological theories of judgement were also explored (Crisp, *in submission*).

The programme of research has now investigated marking in GCSE and A-levels in a number of subjects, for a range of question types and from a number of different perspectives. The diagram in Figure 1 summarises the influences and processes that have been identified as important to the marking process from the research conducted so far. The footnotes indicate which papers report on findings in each area.

Whilst the Core Research team have now contributed significantly to an understanding of GCSE and A-level marking, our investigations are ongoing. As part of some doctoral research, the process of marking coursework is being investigated from a socio-cognitive angle. In *Marking Expertise 2* (Suto and Nádás, 2007a), we are continuing to explore the associations between marking accuracy, apparent marking strategy complexity and question features, this time within the context of GCSE business studies and IGCSE biology marking. In other work the judgement processes involved in moderating vocational portfolios (Johnson and Greateorex, 2007) and grading decisions are being explored.

References

- Barritt, L., Stock, P. L. & Clark, F. (1986). Researching practice: evaluating student essays, *College Composition and Communication*, **37**, 315–327.
- Crisp, V. (2007a). *Comparing the decision-making processes involved in marking between examiners and between different types of examination questions*. A paper presented at the British Educational Research Association Annual Conference, September 2007, Institute of Education, London.
- Crisp, V. (2007b). *Do assessors pay attention to appropriate features of student work when making assessment judgements?* A paper presented at the International Association for Educational Assessment Annual Conference, September 2007, Baku, Azerbaijan.
- Crisp, V. (2007c). Researching the judgement processes involved in A-level marking. *Research Matters: A Cambridge Assessment Publication*, **4**, 13–18. This article summarises key findings from Crisp (2007a; *in press; in submission*).
- Crisp, V. (*in press*). Exploring the nature of examiner thinking during the process of examination marking. *Cambridge Journal of Education*.
- Crisp, V. (*in submission*). A tentative model of the judgement processes involved in examination marking.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, **7**, 31–51.
- Greateorex, J. (2006). *Do examiners' approaches to marking change between when they first begin marking and when they have marked many scripts?* Paper presented at the annual conference of the British Educational Research Association, 6–9 September, University of Warwick.
- Greateorex, J. (2007). Did examiners' marking strategies change as they marked more scripts? *Research Matters: A Cambridge Assessment Publication*, **4**, 6–13. This article summarises some key findings from Greateorex (2006).
- Greateorex, J. & Bell, J.F. (*in press*). What makes AS marking reliable? An experiment with some stages from the standardisation process. *Research Papers in Education*.
- Greateorex J., Nádás R., Suto W.M.I., & Bell J.F. (2007). *Exploring how the cognitive strategies used to mark examination questions relate to the efficacy of examiner training*. Paper presented at the European Conference on Educational Research, 19–22 September, Ghent, Belgium.
- Greateorex, J. & Suto, W.M.I. (2005). *What goes through a marker's mind? Gaining theoretical insights into the A-level and GCSE marking processes*. Paper presented at the 6th annual conference of the Association for Educational Assessment-Europe, November 2005, Dublin, Republic of Ireland.
- Greateorex, J. & Suto, W.M.I. (2006). *An empirical exploration of human judgement in the marking of school examinations*. Paper presented at the 32nd annual conference of the International Association for Educational Assessment, 21–26 May, Singapore.
- Greateorex, J. & Suto, W.M.I. (*in submission*) *What do examiners think of 'thinking aloud'? Interesting findings from an exploratory study*. A version of this paper was presented at the annual conference of the British Educational Research Association, 6–9 September 2006, University of Warwick.
- Johnson, M. & Greateorex, J. (2007). *Assessors' holistic judgements about borderline performances: some influencing factors?* Paper presented at the annual conference of the British Educational Research Association, 5–8 September, University of London.
- 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.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, **19**, 246–276.
- Milanovic, M., Saville, N. & Shuhong, S. (1996). A study of the decision making behaviour of composition-markers. In: M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment*. Cambridge: Cambridge University Press.

- Nádas, R. & Suto, W.M.I. (2007). *An exploration of self-confidence and insight into marking accuracy among GCSE maths and physics markers*. Paper presented at the annual conferences of the British Educational Research Association, 5–8 September, London; and the International Association for Educational Assessment, 17–21 September, Baku, Azerbaijan.
- Pula, J. J. & Huot, B. (1993). A model of background influences on holistic raters. In: M. M. Williamson & B. Huot (Eds.), *Validating holistic scoring for writing assessment: theoretical and empirical foundations*. Cresskill, NJ: Hampton.
- Sanderson, P. J. (2001). *Language and differentiation in Examining at A Level*. PhD Thesis. University of Leeds, Leeds.
- Suto, W.M.I. & Greateorex, J. (2006). A cognitive psychological exploration of the GCSE marking process. *Research Matters: A Cambridge Assessment Publication* 2, 7–11. This article summarises key findings from Suto, W.M.I. & Greateorex, J. (*in press*).
- Suto, W.M.I. & Greateorex, J. (*in press*, a). What goes through an examiner's mind? Using verbal protocols to gain insights into the GCSE marking process. *British Educational Research Journal*. Findings from this paper were also presented at the annual conference of the British Educational Research Association, September 2005, University of Glamorgan.
- Suto, W.M.I. & Greateorex, J. (*in press*, b). A quantitative analysis of cognitive strategy usage in the marking of two GCSE examinations. *Assessment in Education: Principles, Policies and Practice*. Findings from this paper were also presented at the annual conference of the Association for Educational Assessment -Europe, November 2005, Dublin, Ireland.
- Suto, W.M.I. & Nádas, R. (2007a). The 'Marking Expertise' projects: Empirical investigations of some popular assumptions. *Research Matters: A Cambridge Assessment Publication*, 4, 2–5. Paper also presented at the annual conference of the International Association for Educational Assessment, 17–21 September, Baku, Azerbaijan.
- Suto, W.M.I. & Nádas, R. (2007b). *What makes some GCSE examination questions harder to mark accurately than others? An exploration of question features related to accuracy*. Paper presented at the annual conference of the British Educational Research Association, 5–8 September, London.
- Suto, W.M.I. & Nádas, R. (*in press*). What determines GCSE marking accuracy? An exploration of expertise among maths and physics markers. *Research Papers in Education*.
- Vaughan, C. (1991). Holistic assessment: What goes on in the rater's mind? In: L.Hamp-Lyons (Ed.), *Assessing Second Language Writing in Academic Contexts*. Norwood, N.J.: Ablex Publishing Corporation.

An exploration of self-confidence and insight into marking accuracy among GCSE maths and physics markers

Rita Nádas and Dr Irenka Suto Research Division

Background

Introduction

A considerable volume of literature in education and occupational research investigates issues in self-confidence and insight, ranging from college students' post-diction self-assessment (e.g. Maki, 1998; Koch, 2001) to work-related self-assessment (Dunning, Heath and Suls, 2004). However, GCSE markers' perceptions of their marking performance and their metacognition have not, to our knowledge, been examined.

Exploring markers' perceptions is important for several reasons. First, if markers' estimates of their own performance prove to be accurate, then this information could be used by Awarding Bodies in standardisation procedures¹ to identify and discuss examination questions that markers have difficulties with. If, however, markers' insight proves to be unreliable and unrelated to their actual marking accuracy, then their feedback on 'problem areas' could be misleading: for example, when conducting standardisation procedures, Principal Examiners might find themselves focussing on the 'wrong' questions. Secondly, investigating whether self-confidence and insight change or become more accurate with more marking practice or more feedback could inform the marker training practices of Awarding Bodies. This may thereby enhance marking accuracy: there is evidence that improvement of one's self-assessment or

insight into performance results in enhanced test performance (Koch, 2001; Dunning, Johnson, Ehrlinger and Kruger, 2003).

In this article we present the aims and findings of research which explored GCSE markers' perception of their own marking performance, namely, marking accuracy. Markers' levels of self-confidence and insight and possible changes in these measures over the course of the marking process were investigated. The term 'self-confidence' here denotes markers' post-marking estimates of how accurately they thought they had marked a sample of questions; 'insight' refers to the relationship between markers' actual marking accuracy and estimated accuracy, indicating how precise their estimates were.

Theories of insight and self-confidence

Insight into performance has been widely researched from various angles; and it has generally been found that people tend to have incorrect estimations of their own performance. For example, Dunning *et al.* (2003) found that when asked to predict their mastery on an examination, students in the bottom quartile greatly overestimated their actual performance. They also found that the better performing students were able to predict their raw scores with more accuracy, with top performers actually slightly underestimating their scores.

Several theories have been proposed to explain the phenomenon of poor insight. The nature of self-confidence has been examined by cognitive psychologists, who have adopted the 'self-serving bias' theory. Researchers have found that biases are used by participants in research

1 For regulations on standardisation procedures, see Qualifications and Curriculum Authority, 2006