
A qualitative study of the response of raters towards onscreen and paper-based marking

Peter Falvey and David Coniam

The Chinese University of Hong Kong

Abstract

In 2012, onscreen marking (OSM) will replace paper-based marking (PBM) throughout the entire national examinations system in Hong Kong. To chart and validate this major change, a series of studies, mainly quantitative, are being conducted. To complement the quantitative studies this paper reports on an interview-based qualitative study into the responses of raters to OSM and PBM in the English language examination. It investigates the in-depth responses not only of those who have experience of OSM but those who formerly had marked only on paper. Using semi-structured interviews to probe participants' views, issues arose that had not been revealed in the quantitative studies. In addition to revealing abundant insights, new issues were raised in a number of areas, especially reading onscreen; training and standardisation; attitudes towards marking at centres and marking at home; and the accuracy and reliability of marking on screen.

Introduction

Paper-based marking (PBM) will be phased out in Hong Kong public examinations in 2012 and replaced by onscreen marking (OSM). To investigate – and validate – the adoption of OSM in the context of one of the Year 11 (Secondary 5) English language public examinations, for which OSM was adopted as the sole method of marking in a 2007 pilot, a series of studies have been conducted to compare the two modes of marking. The first, Coniam (2009), presents a quantitative examination of ratings for the English language writing examination. The study reported here is an interview-based qualitative investigation that complements the quantitative questionnaire study.

Recent IT developments have allowed OSM to be implemented (see

e.g., Kurzweil, 2005) by public examination bodies around the world e.g., 20 provinces in Mainland China have been using OSM for a number of years and Cambridge Assessment UK have invested in OSM up to 2012 (Raikes, Greatorex & Shaw, 2004). However, at Cambridge, Shaw (2008), noting the growing adoption of OSM in short answer assessments, cautions that there is a paucity of research on the marking of extended responses. Adams (2005), in the United Kingdom's *Assessment and Qualifications Alliance* (<http://www.aqa.org.uk>), voices possible concerns about the validity and reliability of assessments made through OSM. While agreeing that current evidence supports OSM, Adams feels a 'cautious approach' is needed to ensure that stakeholders, in particular governments and teachers, are comfortable with OSM. Notwithstanding this, results of recent studies demonstrate greater comparability between PBM and OSM in terms of the reliability of the results between the two modes of marking (Fowles, 2008; Johnson, Nádas & Bell, 2010; Coniam, 2009, 2010; Coniam & Yeung, 2010).

In 2005, the Hong Kong Examinations and Assessment Authority (HKEAA) established three special OSM centres (a fourth to be established in 2010) with over 1,000 marking-dedicated workstations ready for wholesale implementation of OSM in 2012.

The implementation of OSM has been the incentive for the current series of validation studies – particularly given its implications for other regions and countries where conditions may vary in terms of raters' willingness to change, available resources, and the availability of IT-sophisticated workforces. Partial implementation of OSM has already occurred in the UK (see http://www.rmassessment.co.uk/news/markings/cambridge_2010), Northern Ireland (see <http://www.rewardinglearning.org.uk/newsroom/2010/030810.asp>), New South Wales, Australia where 10% of its public examinations are marked by OSM (see <http://news.boardofstudies.nsw.edu.au/index.cfm/2010/3/22/onscreen-marking>), Cyprus (see <http://www.moec.gov.cy/ypexams/index.html>) and China, where, even with only partial implementation, as many as 3,000 raters took part in the OSM marking of the *gao kao* (China's university entrance examination) in one province alone (see http://www.pxdgc.com/new_1353.html). The

implementation of OSM for the whole public examination system in Hong Kong thus has implications for all jurisdictions where only partial or no OSM rating takes place.

It should be noted that most PBM/OSM comparative studies investigate experienced PBM raters shifting to OSM. This study is different; it investigates the ratings and attitudes of experienced PBM raters moving to OSM as well as raters who have only ever marked via OSM. The first study to follow this procedure - of investigating the ratings and attitudes of experienced PBM raters moving to OSM as well as raters who have only ever rated via OSM - was a quantitative study (Coniam 2009), comparing OSM and PBM in the 2007 Writing Paper of the Hong Kong Certificate of Education (HKCE) English language examination (candidature 99,771). It was a precursor to the main study of this paper. For both studies, efforts were made to recruit first-time raters who had only rated on screen to provide, as stated above, a useful comparison between experienced raters who had always marked on paper and raters who would only have rated on screen.

The Coniam (2009) study showed that statistically, the two rating conditions were similar, with figures for the overall examination indicating that the rating of the sets of scripts marked under the two rating conditions by a single rater compared favourably with previous ratings of the same scripts by two different raters rating scripts. Furthermore, discrepancy rates (5 marks difference out of 24) were as similar between the two mediums of marking as they were between two different raters.

Raters were asked if they could be interviewed to probe, in greater depth, attitudes to OSM and PBM to bring to light issues which may not have been addressed – or even surfaced – in previous quantitative studies involving questionnaires. It would appear that no in-depth qualitative studies have been conducted to probe raters' attitudes. Those who replied positively form the database for the current study.

There now follows a brief description of Hong Kong's education and examination systems, the background to OSM and the OSM process

in Hong Kong, comments on raters' attitudes to OSM and, finally, an account of the main study.

The Hong Kong Education and Examination System

Hong Kong currently operates a system of six primary plus five secondary forms followed by two further secondary years (13 years) – with three bands of ability each covering 33% of the student ability range at secondary level. At present, Hong Kong has two major public examinations conducted by the Hong Kong Examinations and Assessment Authority (HKEAA) at the end of Secondary 5 (Year 11) and Secondary 7 (Year 13).¹

Background to OSM

This section describes marking practices and processes in Hong Kong – in particular the forthcoming OSM - and examines raters' attitudes towards OSM, given the previous tradition of PBM.

The advantages of OSM are claimed to be improved security, enhanced quality assurance (see e.g., Raikes et al., 2004) and increased fairness to candidates when discrepancy scoring occurs between two raters because scripts can be flagged immediately and a third rater is triggered. The disadvantages are the resistance to OSM by experienced PBM raters and the inconvenience of having to travel to designated marking centres at specific times.

¹ From 2009, the secondary school system will operate on a 3+3 model (similar to the Chinese and Australian education models) with a new single examination at the end of Year 12 – the Hong Kong Diploma in Secondary Education (HKDSE). This examination will first be administered in 2012, with undergraduate education subsequently expanding from three years to four.

The Onscreen Marking Process in Hong Kong

In Hong Kong, completed scripts are first delivered to OSM centres for scanning. Raters then attend one of three OSM centres to mark via a secure intranet network (a fourth centre is planned for 2010). OSM was first used in Hong Kong in 2007 when all Year 11 Chinese Language and English Language public examination scripts were marked on screen. This was extended in 2009 to include Liberal Studies.

Once selected, raters first mark trial scripts at home. These are discussed at the subsequent Markers' Meeting to standardize raters' performance. Marking then switches to onscreen mode. Here, raters attempt training scripts by OSM, after which they rate a number of 'qualifying' scripts. These are pre-selected, standardized scripts that assess how well the raters are performing. They must be assessed satisfactorily before raters can begin rating live scripts. If performance is unsatisfactory, raters assess further qualifying scripts. During live marking, control scripts appear at regular intervals (approximately every 50 scripts) to provide instant electronic monitoring, standardisation and maintenance of standards.

While the HKEAA had conducted some small-scale in-house studies of rating via OSM, the Coniam (2009) quantitative study was the first formal study of the efficacy of marking using the two media.

Raters' Attitudes to OSM

The majority of studies, especially more recent studies, generally suggest that the two methods are largely comparable (see Fowles, 2008; Johnson et al., 2010; Coniam, 2009, 2010; Coniam & Yeung, 2010).

However, it is important to note several instances of problems with OSM. In Powers, Farnum, Grant and Kubota (1997), the OSM system was received "relatively positively" (p. 10) by most raters, despite some misgivings. In Zhang, Powers, Wright and Morgan's (2003) study, raters complained of few opportunities to discuss issues with

other raters; the lack of printed commentaries on training essays; and having to scroll down in order to read some essays. Some complained about handwriting image displays and problems connecting to the website. Twing, Nichols and Harrison (2003) reported anxiety about interacting with computers and marking on screen. Adams's (2005) too reported a mixed reaction. It is possible that with the increased adoption of OSM, negative attitudes are diminishing.

The Current Study

The current study focuses on the findings from the detailed responses of raters in semi-structured interviews and explores the issues that emerged in the qualitative data.

This section describes the raters, how their interviews were standardised, the collection of data and the methods used for data analysis. 17 raters were interviewed. These comprised five new raters, and 12 experienced raters.

Data

Appendix 1 provides details of the 17 raters. The five new raters are displayed first. Two of the new raters had less than five years teaching experience. The remaining three had 11-20 years experience. All raters had taught at different levels in secondary school.

While all the new raters were female and three of the 12 experienced raters were male, all are referred to as 'she' for purposes of anonymity. Each rater is referred to by the letter 'N' denoting new rater or 'E' denoting an experienced rater.

Semi-Structured Interviews - Standardisation of Interviewers

Based on the responses to the questionnaire and the open comments that were submitted for the Coniam (2009) quantitative study and in order to facilitate the interviews, the research team drafted the first version of the interview guide focusing on areas of interest that had emerged from the earlier study. In terms of the use of interviews in

qualitative research, Miles and Huberman (1984:21), early commentators on qualitative research methods, note the use of interviews as one of four major methods of qualitative data collection viz. observation, interviews, extracts from documents, and tape recordings, while Silverman (2000) states that interviews are one of the “gold standards” of qualitative research and Harrell and Bradley (2007), in a manual prepared for the US Government, provide extensive explanations of the advantages and disadvantages of semi-structured interviews. Hannan (2007:2) states “They (semi-structured interviews) are the form most often used in education research”.

The training and standardisation of interviewers followed a typical iterative framework by first establishing guidelines with the questions initially prompted by responses to 27 written open-ended comments, elicited in the quantitative study (see Appendix 2 for a list of guideline questions/prompts), then piloting a semi-structured interview which allowed the guidelines to be refined. The interview was transcribed by each of the three interviewers and their scripts were compared for purposes of triangulation and reliability. While this procedure is unnecessary for skilled, experienced interviewers, it proved useful as a training tool for these interviewers in helping them become aware of the process of transcription. In addition, the three interviewers were encouraged to reflect on their interviews and complete a reflective log which they shared with the other two interviewers. The interviewees asserted that this process made them better interviewers. They said it made them more aware of what they were doing and how they were doing it.

Analysis of the interviews was conducted by the author and validated by using other experts to analyse sections of the responses in order to ensure that the analysis was an accurate reflection of the respondents’ views. The process was as follows: watching/listening to and re-listening to the interviews; note-taking and close study of the interviews and the interviewers’ notes; reading and re-reading of transcripts, establishing preliminary categories for analysis based on the areas outlined in the guide for the interviews; creating summaries of interviews; and tabulating responses in those categories. It was then possible to study the categorized and tabulated responses,

describe the results including illustrative quotations and draw conclusions from them.

One of the main advantages of using semi-structured interviews is that although they do not reveal as much of the whole picture as open interviews, they provide an initial framework for categories and the recording of salient details in the subsequent analysis, thus saving time while still allowing the interviewee to feel free to say whatever concerns them. To repeat what Hannan (2007) says: *They (semi-structured interviews) are the form most often used in education research.*

Results and Discussion

This section discusses rater attitudes. The first topic is IT proficiency and the ability to cope with the technical demands of OSM. This is followed by five other issues that arose from the interviews: Reading on screen; Views on the reliability and efficiency of marking by OSM; Training and standardisation; Attitudes towards OSM marking centres; and Views on working from home.

IT Proficiency and the Ability to Cope with the Technical Demands of OSM

All seventeen raters felt that they had sufficient levels of IT proficiency to cope with the demands of OSM. This, unlike earlier findings from other countries, appears not to be an issue either for raters or for the HKEAA. All confirmed in their interview that the skills they possessed or the skills that they were given, were adequate for the task.

One new rater, N1, had initial difficulties but was able to overcome them satisfactorily:

... on the first day of marking I feel quite helpless and I always call for help from the technicians. They are quite helpful. Very soon, I can get used to it. (N1)

One experienced rater, E8, the oldest in the group, said she was a rare

user of computers but was able to cope well after she had been briefed in training:

The marking difficulties only occurred in the early process of the marking process because, as I have mentioned, I am not good at computer so my difficulties occurred. But after that particular the briefing session, I understood a lot of important (things) up to now, everything goes smoothly. (E8)

The concerns expressed in other jurisdictions in previous studies did not manifest themselves in this study either for new or experienced raters. It appears to be a non-issue in Hong Kong and, as computer literacy grows in other countries it is likely that this will diminish as an issue.

Reading On Screen

No major issues arose in the Coniam (2009) questionnaire-elicited data on this topic but the interview data revealed a major issue – eyes became tired when reading on screen. N2 said that she found onscreen reading ‘a bit tiring’ but that was because she had booked three hours at the marking centre and did not like to take too many breaks because she wanted to finish her allocation of scripts. Although she originally said tiredness of the eyes was not a problem for her, commenting on the experience, she rather contradicts herself, saying:

It's a bit tiring reading scripts on the screen. But I think it's more convenient for me to mark a lot of scripts because I don't need to flick the pages. I don't need to carry the papers from here to there. So I think it's good for me to mark on screen. (N2)

Two of the five new raters experienced problems. Seven experienced raters said OSM tired their eyes. In addition, E9 said that not only were her eyes tired, she also felt dizzy after a long stretch at the screen while N17 said that although her eyes were not as tired as she had anticipated, her neck became sore. OSM is a new experience for raters and for the HKEAA. It is clear that even new raters may feel

some tiredness while many experienced raters complain that their eyes become tired when reading on screen. These complaints may decrease with time as raters adjust to onscreen marking. Indeed, it is worthy of note that within the general context of reading on screen, the news that Apple's Ipad is expected to ship 63.3 million units in 2011 (<http://www.macdailynews.com/>) and the fact that Amazon's Kindle reading tablet (<http://www.amazon.co.uk/>) was its biggest-selling product in 2010, would appear to indicate that onscreen reading is becoming more and more acceptable.

A further issue arising from the qualitative data involved problems experienced by raters with the legibility of scanned scripts when test takers used correcting fluid. N5 says:

Sometimes they are well-scanned. But sometimes when the candidates use the correction pen to correct what they have written, then there may be some problem in scanning. (N5)

E7, also commenting on the use of correction fluid states:

It's okay. But if there are some papers with correction pens, the letters, the words are blurred. So, this will make it very hard to determine whether students did make mistakes. (E7)

The issue of legibility may have to be addressed more seriously if scanned scripts become genuinely illegible. The problem will diminish as more powerful scanners are developed but, for the moment, illegible scripts remain a problem especially among raters, whether new or experienced, who complain of eye strain - even though Coniam (2010) makes it clear that equipment is of a high standard with dedicated workstations provided with 21" rotatable monitors.

Views on the Reliability and Efficiency of Marking by OSM

The qualitative data revealed a number of insights into raters' views on the reliability and efficiency of marking by OSM. As would be expected, with the implementation of an innovative rating practice,

raters revealed variable and conflicting views on issues of reliability and efficiency when using OSM. E7, for example, a highly experienced teacher and rater, was strongly against OSM but admitted that it did have specific advantages. Commenting on reliability, she makes two specific comments in which she says that the use of OSM will increase reliability:

Technically speaking, it's good. It really increased the reliability of the whole assessment. And, I have to admit that this increases the transparency of sharing statistics among senior markers. (E7)

On the one hand, this (the use of OSM) will increase reliability as well. Because as far as I know, it's easier for the senior marker to understand the progress and the problems some junior markers are facing. (E7)

However, in spite of increased reliability, E7 feels that her accuracy suffers using OSM:

I could quite confidently tell you that my speed of marking on screen is surely lower than marking live scripts. If you further ask me if this could help the teachers mark the essays accurately, I would say not really especially after having marked for a long time. Because, usually markers have stayed in the marking centre for 3 hours. So, the possible result is that some of the scripts being marked at the end, the standardisation would turn out to be not that accurate. (E7)

Marking efficiency was characterized by the raters as the ability to get through more scripts in one mode or the other without compromising accuracy or reliability. All new raters except N5 said they were more efficient using OSM. N5 said she was equally efficient with both modes.

The experienced raters are less unanimous than the new raters in their attitudes to the efficiency of using OSM: three felt that they were more efficient with OSM in a marking centre, four said that there was no difference in efficiency between the two modes and five stated that they were more efficient using PBM. However, one of the five

produced a caveat that qualified her support for PBM, particularly because of the advantages brought by the use of control scripts. She says:

It's different on the computer, you have the control scripts. So you will compare your script to the other, so you know how accurate you are. It's good. And you know the number; everything is automatically shown to you. You know how many scripts you have marked. So I enjoy those kinds of convenience. (E13)

When the Interviewer asks why, she replies:

... I can retrieve the past papers, I can look at the record and look at the time I spent on each script. At home I think I can mark faster but I cannot enjoy those kinds of conveniences..... It's different on the computer, you have the control scripts. So you will compare your script to the other, so you know how accurate you are. It's good. (E13)

One new rater, N2 who prefers the use of OSM, comments on the ease of access that raters enjoy when browsing through previously marked scripts. This is in marked contrast to rater E10's views in the section below. Rater N2 says:

I think the marking accuracy would be higher for onscreen marking because it's easier for the users to browse through the previous scripts. But if they're marking the paper-based scripts, they won't have the incentive to refer back to the previous scripts. (CKW-N2)

We can thus see that there is a broad spectrum of views on the accuracy and efficiency of using OSM. More new raters than experienced raters felt that they were more efficient with OSM. It will be interesting to investigate whether this spread of views continues over the next few years when the attitudes of raters of English who become experienced in OSM are compared to the attitudes of raters of other subjects who will only start OSM in 2012. In the next short section, one perceived disadvantage of OSM is discussed.

Re-reading of scripts

Although rater N2, above, states 'I think the marking accuracy would be higher for onscreen marking because it's easier for the users to browse through the previous scripts', rater, E10, an experienced rater disagrees:

I feel more comfortable with scripts (on-paper) because I tended to re-read some of the scripts. But for onscreen marking, it's quite difficult... Because the scripts are ordered according to the time you mark it. (E10)

Clearly, Rater E10, who has rated both on paper and on screen prefers the original method for accessing scripts that need to be re-read. The new rater, N2, does not perceive this to be a problem. However, discussions with officials at the HKEAA revealed that the interviewees such as E10 were accurate in their observations. They confirmed that although it is possible to re-call scripts that have been marked on screen, it is much more difficult than sorting through a pile of scripts at home. This issue, discussed further in the Conclusion, will need to be addressed in future by the HKEAA, particularly for experienced raters.

Training and Standardisation

No strong views about training and standardisation were expressed in the quantitative data. However, some significant issues emerged in the qualitative data which may require follow-up. The five new raters stated that the training was either acceptable or good. Two would have preferred more training but they were complimentary about the quality of training (however see issues arising from the use of qualifying and control scripts below). The proportion of experienced raters who found the training experience acceptable or good was less than for the new raters.

Eight experienced raters felt that the training they received was satisfactory. Two, E7 and N15 felt that new raters needed the training more than they did and that it focused on marking training to the

exclusion of technical/operations training while N9, with a contrary view, felt that even more time for marking training was required.

E11 agreed and felt that the training she received was too rushed. She says:

The training is quite rushed because they just give you the script and then they tell you the comments and give marks to you. (E11)

There was some confusion about qualifying scripts. The majority criticised them mainly because of the small number of scripts that were used. A new rater says:

... there are only four qualifying scripts, so I don't think they would be that useful....More would be better. (N2)

E14, one of the many who agrees with N2, says:

Before you start marking, you have to mark those qualifying scripts. There are only four to five, so in a sense I don't think it's enough for me to understand the marking codes and the ways of marking. (E14)

The confusion occurred because additional qualifying scripts – over and above the initial four – are given to raters only if they do not do well enough when rating the initial qualifying scripts. If the initial qualifying scripts are rated satisfactorily, the raters move straight on to live scripts. E17 shows her lack of understanding of the use of qualifying scripts when she says:

Is it possible to have more sample scripts so that we have a better picture of what are the possible criteria we have to consider when it comes to [unclear]. I hope other teachers also reflect the fact that this area is very weak. (E17)

Another concern was that a Senior Examiner, undoubtedly under pressure during the training session, informed raters of the marks that had been awarded to scripts used in training instead of making the raters work through them. New examiner N2 says:

Basically, there were two parts, the first part required us to read it and mark it. The second part required us to sit together and mark another set of scripts and talk with our leader. There was a Senior Examiner as our leader. But the first part, the Senior Examiner didn't give us enough time to read the scripts and he just directly told us the mark. (N2).

Further corroboration is provided by E14:

They give quite a good sample. But I think at the markers' meeting, there may not be enough time for discussions. Usually it's the Senior Examiner telling you which mark should be given for which script, and some very brief explanation. (E14)

This is a serious concern, one that has the potential to affect the reliability of raters and although it may have been a single aberration, the HKEAA may wish to consider further senior examiner/coordinator training to ensure that such behaviour is not repeated in future years.

After training, raters begin rating live scripts. During this phase, control scripts appear at regular intervals to monitor and standardise raters' performance. Because information about rating is provided instantaneously to Senior Examiners, intervention can occur when necessary to ensure that standards are maintained. All five new raters approved of this innovation. One noted that the use of control scripts in OSM could bring pressure to bear on raters but approved, nonetheless, of the procedure:

Yes, in fact, I mean, the coming of control scripts might give me lots of pressure. They are really checking my scoring. On the other hand, it was quite useful to me, to adjust my marking style, my tendency to give high marks, etc. I think the idea of giving control scripts is good. (N1)

A highly experienced rater approves wholeheartedly of the use of control scripts:

Control scripts are interesting and fascinating. Challenging. I can tell you that the most important thing to support my energy, perseverance and concentration is the control scripts. I found the control scripts extremely useful for the examiners or markers to know whether the marks he/she has given is relevant or proportionate to the standard score or not, especially for some new markers. (E8)

However, the qualitative data revealed that raters were quick to note the appearance of control scripts was predictable. They appeared every 50 scripts or so. Many made suggestions that their occurrence should be less predictable. An experienced examiner states:

It's useful but it's predictable. You know what I'm saying? You can predict it. (E6)

Another rater, E14, confirms this:

To be frank with you, I would try and anticipate when the control script will come, after fifty scripts. So, if they want to make it work, they should give them more randomly. (E14)

Further concerns about the use of control scripts emerged during the interviews. One relates to the omission of control scripts after raters have finished their first batch of 1,000 and move on to more scripts. N3, concerned about maintaining standards, explains the problem she faced when asked to continue marking after 1,000 scripts:

.... you have to keep sending some control papers; because I really want to make sure that I will not go too far away from the standard marking. You know, I don't want to be unfair to any student. So this is the problem that I found in my last year's experience. I found that there was no control paper after one thousand. (N3)

This, presumably, is a problem that can be solved relatively simply by the HKEAA. One possibility would be for the distribution system to be re-programmed to provide further control scripts once the 1,000th script has been assessed.

A final issue concerns inconsistencies in the marks awarded to control scripts:

I think the qualifying scripts are quite useful. But sometimes we think the control scripts... there might be some problems on the control scripts, because the marks given are not quite consistent, according to their guidelines. (N5)

Rater E12 agrees:

But just that sometimes the standard is not that consistent in some areas. (Interviewer: The standard of marking?) Yes, the standard of marking. Sometimes I think this can be assigned a higher mark, but actually the mark is different from what I expect. I think the HKEAA people, the chief and senior examiners; they need to have a better compromise. (E12)

Issues of this nature tend to occur in most examination contexts. To address them, it is necessary to monitor the situation, and allow those who have problems with the grades awarded by the 'experts' to discuss them more fully with Senior Examiners. In this way, inconsistent 'expert' marks can be adjusted and the adjustment made through the software to reflect this. In addition, Senior Examiners should attempt to ensure that the time spent selecting and grading sample scripts for control purposes is fruitful. This ability to flag up problems quickly is one of the advantages of OSM and its use of control scripts. Control scripts in OSM now form an essential part of the training and standardisation process and are necessary for purposes of reliability and validity.

Attitudes Towards OSM Marking Centres

Many of the findings from the quantitative data were replicated in the qualitative data e.g., new raters had fewer problems with the marking centres than did the experienced raters and in terms of the location of the marking centres for OSM, the raters' responses were mainly dependent on how close they lived to the nearest centre. Those who lived far away from a centre, particularly those who came

from the New Territories (a long way from Hong Kong Island) found travelling inconvenient. Rater E14 states:

I have to spend 45 minutes traveling to the Tsuen Wan Centre. It's the time that counts ...and the repetitiveness. Actually if I could mark it on paper, I can arrange my own time. (E14)

There were some surprises, however, in the qualitative data. Raters' views of the environment, facilities and ambience of the marking centres showed a distinct preference for the Wan Chai centre on Hong Kong Island. A number of respondents said it was 'lighter, airier, and more spacious than the others'. N2, for example, reported that she travelled a long way to get to the Wan Chai centre even though she lived in Tuen Mun (at the opposite end of Hong Kong):

It's quite tiring for me because I live in Tuen Mun. The nearest marking centre is in Tsuen Wan. But I hate that. So basically I go to Wan Chai every day. Every night. So it takes a long time on traveling. (N2)

Views on Working from Home

New insights were provided by the qualitative data. While the quantitative data broadly revealed that the new raters were satisfied with the marking centres, there was a clear exception. N3 is typical of most of the new raters. She says:

Yep, I think it's more convenient than I have to find out the papers at home because you know the pile of papers is that thick. You take it home, and you have to search it and read it again. But if you use the online... onscreen marking system you can click on the date, and then you will see all the scripts in order. So I think it's easier for you to... to adjust the marks if you want (N3).

The exception is N5, a new rater and an experienced teacher. She would prefer to do her marking on paper at home. While she is IT proficient, happy with the technical operation of OSM, can read well off the screen, and is content with the location and physical facilities

at the marking centre, she is used to PBM and does not want to change. The finding is not surprising as research shows that teachers prefer change to be gradual. They do not like having to change their beliefs substantially (Nespor, 1987; Borko and Putnam, 1996). Wall (2005), citing research done with teachers in the field of assessment notes that often what many instigators of innovation and educational reform consider minor, is perceived by other stakeholders (in this case raters) as major. N5, having been used to marking homework, school examinations and papers for other examination boards at home, found the burden of having to leave home for a marking centre at fixed times disruptive and inconvenient. A further slightly surprising finding is that of two of the new raters who prefer OSM would nonetheless rather mark at home. Marking at home clearly has many attractions. Of the two latter raters, N2 stated that she would prefer OSM at home but accepts that problems of confidentiality and security would preclude it being introduced. It should be noted, however, that OSM rating from home is practiced in other jurisdictions, e.g., in Australia, the New South Wales Government Board of Studies reports on a system of onscreen marking introduced in 2009 to enable raters to mark from home via a secure Internet website (see <http://news.boardofstudies.nsw.edu.au/index.cfm/2010/3/22/onscreen-marking>).

Of the experienced raters, all but one, E9, would prefer to mark at home whether it be PBM or OSM. E9 says she would prefer PBM at a marking centre because:

I know that in some places like Australia, markers would have to stay in a place and then they mark the actual script. Then I think this is better because you don't have to look at the computer screen.
(E9)

Some of the experienced raters who would rather rate at home accept that concerns of security and confidentiality would cause problems for OSM (e.g., E14) so there is an awareness of constraints on their preferred place of marking. As stated above, however, other jurisdictions accept OSM home marking without concern.

Talking about using the OSM centres, E16 says:

So there were people moving around. And they come to your station and they log on. And they move around and take a break. But like if I work at home then I just lock myself in my room when my child is sleeping. And no TV... nothing. I can concentrate in my work. (E16)

E1, another supporter of marking at home worries that experienced raters might be lost to the system because they can no longer mark at home.

I am negative towards the system. But I get that the EA (Examinations Authority) is not moving towards an easier system, they bought so many new computers. I think they have sacrificed some of the experienced markers (E17)

This is an issue which, in time, should diminish although, in the meantime, HKEAA should do its best to make the OSM experience as rewarding as possible to mitigate the dislike of having to attend designated centres.

Conclusion

The qualitative study has undoubtedly provided new insights into the attitudes and behaviours of the raters over and above the results of the quantitative study. The evidence above reveals evidence of some disquiet about using the marking centres for extended periods, a considerable number of complaints about travelling, a marked preference for the convenience of marking from home, observations about the predictability of the appearance of control scripts, concerns over the way in which supervisors indicated what the scores were on the qualifying scripts and problems with the legibility of scripts on screen.

Another issue focused on individual raters accessing scripts for re-marking, which is easier at home than in a marking centre. Discussions with officials at the HKEAA revealed that the interviewees were accurate in their opinions in that, although it is

possible to re-call scripts that have been marked in the Marking Centre, it is much more difficult than sorting through a pile of scripts at home. As OSM is to be universally adopted in Hong Kong, this issue will need to be addressed, possibly through re-writing part of the software to allow easier access to scripts that have been previously marked.

On a positive note, most interviewees felt that security was strengthened by the use of OSM. Press reports of past incidents of lost scripts have impinged on the Hong Kong public's awareness and all respondents realized that OSM presents a level of heightened security not previously available with PBM.

What is clear is that the interviews revealed issues and concerns that should be investigated in future. There should be follow-up studies for English and Chinese examinations to investigate whether the findings of the qualitative and quantitative studies are replicated and, in addition, whether improvements to training and sheer familiarity with OSM reduce complaints. Practical solutions can be suggested such as: ensuring that the best scanners are regularly purchased and installed to help legibility; inaugurating a campaign to reduce the amount of correcting fluid used by students to improve raters' efforts to read scripts; ensuring that the best equipment is always used e.g., size and quality of monitors; and making strenuous efforts to persuade experienced raters that the new systems are, overall, of benefit to them as raters and to the students who have written the scripts.

Furthermore, it will be necessary for future validation studies to be conducted once OSM becomes universally implemented in 2012; e.g., how raters of other subjects rate using OSM and how they react to it.

References

- Adams, C. (2005). *How does assessment differ when e-marking replaces paper-based marking?* Paper presented at the 31st Annual International Association for Educational Assessment Conference "Assessment and the Future of Schooling and Learning". September 4-9 2005, Abuja, Nigeria.
- Amazon Kindle reading tablet. Retrieved January 12 2011 from http://www.amazon.co.uk/dp/B002LVUWFE/?tag=yahhyd-21&hvadid=68244811531&ref=pd_sl_xpsdjopo3_b.
- Apple I-pad. Retrieved 11 January 2011 from http://www.macdailynews.com/index.php/weblog/comments/suppli_ups_ipad_forecast_to_63.3_million_units_in_2012/.
- Borko, H., & Putman, R.T. (1996). Learning to teach. In Berliner, D.C., & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp.673-708). New York: Macmillan.
- Cambridge Assessment. (2010). Retrieved 11/1/2011 from http://www.rmassessment.co.uk/news/marketing/cambridge_2010.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Coniam, D. (2010). Validating onscreen marking in Hong Kong. *Asia Pacific Education Review*, 11(3), 423-431.
- Coniam, D. (1991). Essay marking: a comparison of criterion-referenced and norm-referenced marking. *Institute of Language in Education Journal*, 7, 154-164.
- Coniam, D. & Yeung, A. (2010). Markers' perceptions regarding the onscreen marking of Liberal Studies in the Hong Kong public examination system. *Asia Pacific Journal of Education*, 30(3), 249-271.

-
- Council for the Curriculum Examinations and Assessment, Northern Ireland (2010). Retrieved 11 January 2011 from <http://www.rewardinglearning.org.uk/newsroom/2010/030810.asp>.
- Directorate of Higher Education, Ministry of Education and Culture, Cyprus. Retrieved 11 January 2011 from <http://www.moec.gov.cy/ypexams/index.html>.
- Fowles, D. (2008). Does marking images of essays on screen retain marker confidence and reliability? Paper presented at the International Association for Educational Assessment Annual Conference, September 7-12, Cambridge, UK
- Hannan, A. (2007). Interviews in education research. Faculty of Education, University of Plymouth. Retrieved January 11, 2011 from <http://www.edu.plymouth.ac.uk/resined/interviews/inthome.htm>.
- Harrell, M.C. & Bradley, M. A. (2007). Semi-structured interviews and focus groups. RAND, National Defense Research Institute. Retrieved January 1, 2011 from http://www.rand.org/pubs/technical_reports/2009/RAND_TR718.pdf.
- Johnson, M., Nádas, R., & Bell, J.F. (2010). Marking essays on screen: An investigation into the reliability of marking extended subjective texts. *British Journal of Educational Technology*, 41 (5), 814-826.
- Kurzweil, R. (2005). *The singularity is near*. Penguin: London.
- Miles, M. & Huberman, M. (1984). *Qualitative data analysis: A source book for new methods*. Thousand Oaks, CA: Sage Publications.
- New South Wales Government Board of Studies. Retrieved March 22 2010, from <http://news.boardofstudies.nsw.edu.au/index.cfm/2010/3/22/on-screen-marking>.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317-328.

-
- Powers, D.E., Farnum, M., Grant, M., & Kubota, M. (1997). A pilot test of on-line essay scoring (ETS RM-97-07). Princeton, NJ: Educational Testing Service.
- Provincial Education Examination, Zhejiang Province, China. (2010). Retrieved 12 June 2010 from http://www.pxdgc.com/new_1353.html
- Raikes, N., Grotorex, J., & Shaw, S. (2004). From paper to screen: some issues on the way. Paper presented at the IAEA Conference, Manchester. Retrieved May 20 2008, from <http://www.ucl.ac.uk/assessmentdirector/articles/confproceedingsetc/IAEA2000NRJGSS>.
- Shaw, S. (2008) Essay marking on-screen: implications for assessment validity. *E-Learning and Digital Media*, 5 (3), Retrieved January 11, 2011 from http://www.worldwords.co.uk/elea/content/pdfs/5/issue5_3.asp.
- Silverman, D. (2000). Doing qualitative research: A practical handbook. London: Sage.
- Twing, J., Nichols, P. & Harrison, I. (2003). The comparability of paper-based and image-based marking of a high-stakes, large-scale writing assessment. International Association for Educational Assessment Conference, Manchester, October 5-10, 2003.
- Wall, D. (2005) The Impact of High-Stakes Testing on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory. *Studies in Language Testing* Volume 22. Cambridge ESOL and Cambridge University Press.
- Zhang, Y., Powers, D., Wright, W. & Morgan, R. (2003). *Applying the online scoring network (OSN) to advanced program placement program (AP) Tests*. ETS Research Report RR-03-12.

Appendix 1: Interviewees' Profiles

No.	Names	New / Exp'd rater	Age	Sex M / F	Teaching level S1-3 / S4-5 / S6-7			Tchg exp'nce (years)	Panel chair
1.	N1	N	26-30	F	-	Y	-	2-5 yrs	N
2.	N2	N	26-30	F	Y	Y	-	2-5 yrs	N
3.	N3	N	31-40	F	Y	Y	-	11-20 yrs	N
4.	N4	N	31-40	F	Y	Y	-	11-20 yrs	N
5.	N5	N	31-40	F	Y	Y	Y	11-20 yrs	N
6.	E6	E	31-40	F	-	Y	Y	11-20 yrs	N
7.	E7	E	31-40	M	Y	Y	Y	11-20 yrs	Y
8.	E8	E	41-50	M	-	Y	Y	> 20 yrs	N
9.	E9	E	31-40	F	Y	Y	-	11-20 yrs	N
10.	E10	E	41-50	F	-	Y	-	11-20 yrs	N
11.	E11	E	31-40	F	Y	Y	-	11-20 yrs	N
12.	E12	E	31-40	F	-	Y	Y	6-10 yrs	N
13.	E13	E	31-40	F	Y	Y	-	6-10 yrs	N
14.	E14	E	26-30	F	Y	Y	-	6-10 yrs	N
15.	N15	E	31-40	M	Y	Y	-	6-10 yrs	Y
16.	E16	E	31-40	F	Y	Y	Y	6-10 yrs	N
17.	E17	E	41-50	F	Y	Y	Y	11-20 yrs	N
Total:		N – 5 E – 12		M – 3 F – 14					2

Key: Column 3, raters: N= new rater; E= experienced rater

Appendix 2: Checklist of guideline questions for the semi-structured interviews

The checklist begins with a repetition of the research hypotheses:

Hypothesis 1 - raters will be sufficiently competent technologically to accept the new OSM medium

Hypothesis 2 - raters will not be negative in terms of the OSM medium, having no preference for either marking medium

A. Attitudes towards using computer

Are you computer savvy?

How often do you use a computer at work or at leisure?

How do you find the technical operation of the OSM medium?

Is the system difficult or easy to manipulate?

Are you comfortable reading off the screen?

How do you find the ergonomic design (comfortableness and ease of operation) of the computer facilities at the marking centers?

B. Attitudes towards OSM

How do you find OSM?

Are there any advantages to OSM when compared with PBM e.g. reliability and consistency between markers?

Do you find the control scripts and qualifying scripts useful.

Do the control and qualifying scripts help to ensure quality and eliminate the potential for disparity?

Can you tell us about any improvements you would like to see in OSM?