Routledge
Taylor & Francis Group

Educational Research and Evaluation

An International Journal on Theory and Practice

ISSN: 1380-3611 (Print) 1744-4187 (Online) Journal homepage: http://www.tandfonline.com/loi/nere20

A qualitative study of markers' perceptions on onscreen marking in five subject areas

Min Yang, Zi Yan & David Coniam

To cite this article: Min Yang, Zi Yan & David Coniam (2018): A qualitative study of markers' perceptions on onscreen marking in five subject areas, Educational Research and Evaluation, DOI: 10.1080/13803611.2018.1446836

To link to this article: https://doi.org/10.1080/13803611.2018.1446836

Published online: 13 Mar 2018.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=nere20

EDUCATIONAL RESEARCH AND EVALUATION, 2018
https://doi.org/10.1080/13803611.2018.1446836

Routledge
Taylor & Francis Group

Check for updates

A qualitative study of markers' perceptions on onscreen marking in five subject areas

Min Yang, Zi Yan and David Coniam

Department of Curriculum and Instruction, The Education University of Hong Kong, Tai Po, N.T., Hong Kong

ABSTRACT

This paper reports on a qualitative study on markers' perceptions of onscreen marking (OSM) in association with key influential factors of marking reliability. The study has made adaptations to an existing framework proposed by Black, Suto, and Bramley in 2011 for exploring issues related to influential factors of marking reliability in OSM contexts. Specifically, the study investigated those influential factors in the framework that can be manipulated and managed by the examination authority. The study involved 31 markers for a large-scale publication examination who were secondary school teachers from 5 subject areas. The major themes indicated that markers' attitudes towards OSM might be potentially associated with the types of questions they marked. Further, the advantages and disadvantages of OSM as perceived by the markers were likely to be related to key influential factors of reliability pertaining to the quality control process and the usability of technology in the OSM system.

ARTICLE HISTORY
Received 6 June 2017
Accepted 26 February 2018

KEYWORDS
Onscreen marking (OSM); public examinations; marking reliability; quality control; OSM technology

Introduction

This paper examines the perceptions of markers in five subject areas regarding key factors that might influence the reliability of onscreen marking (OSM) in the Hong Kong Diploma of Secondary Education (HKDSE) – the Year 12 public examination for admission of school leavers into higher education institutions. The five subjects included: business, accounting, and financial studies (BAFS); health management and social care (HMSC); geography; biology; and history.

OSM is a form of e-marking widely applied in different countries, such as China, the UK, and the USA (Zhang, Powers, Wright, & Morgan, 2003). Unlike paper-based marking (PBM) that is conducted manually on printed scripts, OSM exploits information technology by scanning scripts into images, storing such images in the OSM environment, and transmitting them to markers through the examination authority's OSM system on the intranet or internet (Haggie, 2008). Because a scanned script can be divided into different parts, it is possible to distribute the same script among different markers, which enables marking at the question level and facilitates double marking – both features have been reported in existing research as contributing to improved marking reliability (Tisi, Whitehouse, Maughan, & Burdett, 2013).

CONTACT Min Yang myang@eduhk.hk
© 2018 Informa UK Limited, trading as Taylor & Francis Group

The use of OSM technology also facilitates quality control through mechanisms such as training and feedback provision for markers in the OSM system, which help markers maintain reliability in marking (Suto, Nadas, & Bell, 2011). Thus, it is unsurprising that a survey conducted by RM Education Plc with 17 examination organisations reported the potential for enhancing marking quality to be the top driver for the widespread uptake of OSM and other forms of e-marking among examination organisations internationally (Haggie, 2008).

In the context of OSM, where marking reliability depends on human markers' agreement in applying assessment standards to examination papers (Baird, Greatorex, & Bell, 2004; Campbell, 2005), there is a need to understand the complex issues surrounding marking reliability. Examples of such issues include how markers undergo quality control processes (e.g., training and monitoring of markers' performance) in the OSM system, interact with the technology-supported OSM features, and build consensual assessment knowledge through discussions with peer markers to attain agreement in marking (Black, Suto, & Bramley, 2011). These issues are related to the influential factors of marking reliability, as indicated by our review of studies on such factors in the context of OSM.

Factors influencing reliability in OSM

Reliability is the accuracy and consistency of marking; it represents the likelihood of a candidate receiving the same mark if the assessment procedure is to be replicated (Bejar, 2011; Newton, 2013). Bramley (2007) differentiates two facets of reliability. For assessment questions with restricted responses (e.g., multiple-choice questions) that require minimum marker judgment, reliability denotes *accuracy* of marking. Where essay-type questions with longer, constructed responses demand marker interpretation of the mark scheme, reliability is better described as *marker agreement*.

Meadows and Billington (2005) identify three groups of factors affecting reliability: (a) factors at the level of question/item and mark scheme; (b) factors pertaining to candidates and their answers; (c) factors related to markers. In a review of 28 studies, Tisi and colleagues (2013) conclude that reliability is generally higher for structured, analytically marked questions, while reliability becomes lower when the questions are of an essay type. Marker agreement is also influenced by whether the mark scheme is objective, points based, or levels based, with reliability decreasing when more sophisticated marker judgment is demanded, which invokes simple or complex marker cognitive strategies (Black, 2010; Suto et al., 2011). Candidates' characteristics (e.g., gender, race) and how they respond to questions were also found to affect markers' reliability (Meadows & Billington, 2005). Marker characteristics induce bias as well. For example, less experienced markers are likely to be more severe and less accurate than experienced markers; more subject knowledge, teaching experience, and marking experience were associated with higher reliability (Black et al., 2011).

Given the various influential factors, mechanisms such as marker training, continuous monitoring, and feedback from the Chief Examiner, and markers' coordination meetings for making decisions on the "definitive mark" based on the mark scheme are implemented by examination organisations, with generally positive effects on reliability (Meadows & Billington, 2005; Suto et al., 2011). These mechanisms are discussed in relation to quality control procedures in the OSM system in the current paper. Quality control is a sub-

process of quality assurance for detecting deviations from quality standards (Bejar, 2011). In OSM environments, quality control procedures are strengthened by using information technology (Haggie, 2008).

In the present study, Black and associates' (2011) framework on the interaction of key factors that influence marking reliability has been adapted for understanding the particular issues arising from markers' experience of engaging in OSM processes, and whether these issues were interrelated with some of the influential factors of marking reliability (see Figure 1). The framework displayed in Figure 1 incorporates the above-examined factors identified in the literature as well as OSM-specific factors identified from the research team's previous studies (Coniam & Yan, 2016; Yan & Coniam, 2013, 2014). In Figure 1, boxes with dashed lines and grey shades represent factors that can be manipulated directly or indirectly by the examination organisation (Type 1 factors), while boxes with solid lines (without grey shades) signify factors that cannot be manipulated (Type 2 factors).

The current paper seeks to contribute to the literature by focusing specifically on *Type 1 factors* in the adapted framework, since these are directly associated with markers' experience, based on qualitative findings in the context of the HKDSE in Hong Kong. Such

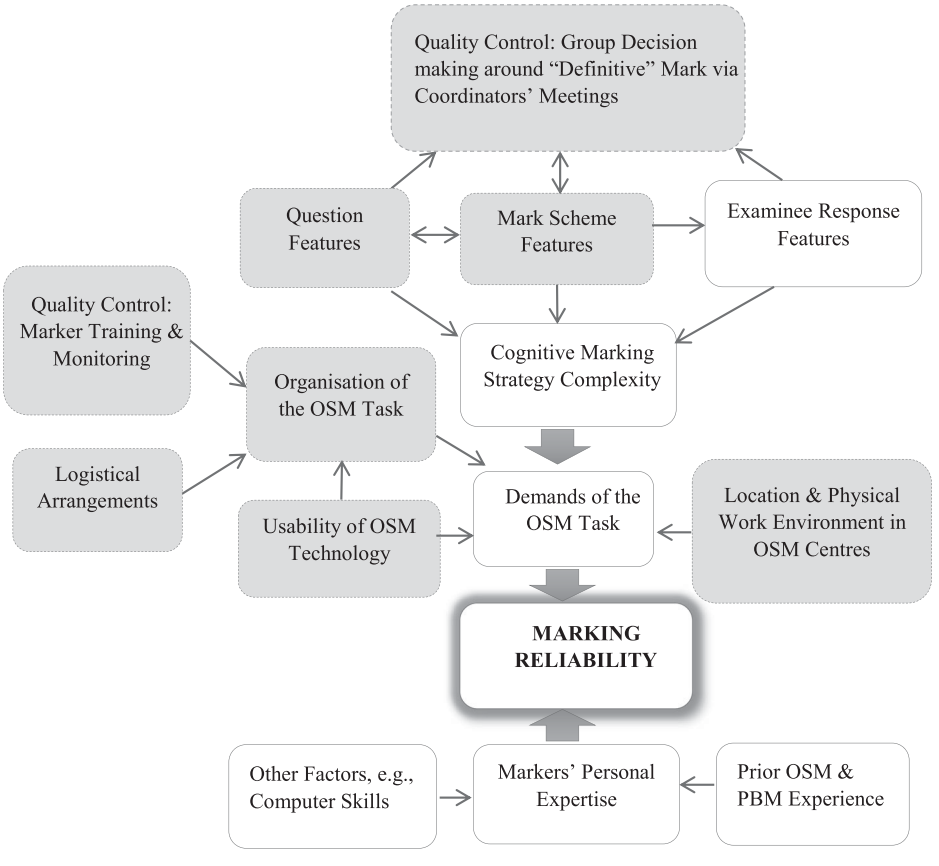

Figure 1. An adapted framework on the interaction of factors influencing marking reliability in OSM (adapted from Black et al., 2011).

qualitative findings provide rich information on how the Type 1 factors can be managed by the examination organisation through formulating strategies for enhancing marking reliability in OSM. Type 2 factors, such as markers' cognitive marking strategy complexity and examinees' response features, are beyond the scope of the current paper.

To fulfil the above-mentioned purpose of this paper, three adjustments are made in the adapted framework on factors influencing reliability in OSM, as compared to Black and colleagues' (2011) original framework. First, the factor "Organisation of the OSM Task" is specified as consisting of two components (i.e., quality control mechanisms – marker training and monitoring, and logistical arrangements related to OSM), which is necessary for detailed analysis of marker experience of the OSM process. Second, the phrase "Physical work environment" is replaced by "Location & Physical Work Environment in OSM Centres" in order to explore whether and how the location of OSM centres influenced markers' experience. Third, the factors "Maximum Mark Tariff" and "Intended Difficulty for Examinees", which are specific to the factors of "Question Features" and "Mark Scheme Features", respectively, are omitted from the adapted framework, since these factors are beyond the scope of the present study.

In Hong Kong, the HKDSE has attracted much public attention as a key element of major curriculum reform implemented in 2012. Under the new curriculum, secondary education has been reduced from 7 years to 6 years, with an additional year added to the previous 3-year undergraduate education. Parallel with this major curriculum shift was the replacement of two previous public examinations (i.e., the Hong Kong Certificate of Education or HKCEE, and the Hong Kong Advanced Level Examinations or HKALE) by a single public examination – the HKDSE, which takes place at the end of Secondary 6 (Year 12). All scripts of the HKDSE have been marked through OSM since 2012, with an annual candidature in the region of 70,000. As an innovation implemented by the Hong Kong Examination and Assessment Authority (HKEAA), the OSM system has been subjected to validation studies on its comparability with PBM regarding reliability and validity (Lee, 2016).

In a series of three studies (Coniam & Yan, 2016; Yan & Coniam, 2013, 2014) conducted by the research team, a two-scale questionnaire was employed for validating OSM in the HKDSE concerning markers' attitudes towards OSM: (a) perceived ease of use in the OSM environment and (b) acceptance of OSM. The first scale tapped into markers' perception regarding their computer skills, usability of OSM technology, and the extent of comfort felt when marking in OSM centres. The second scale examined markers' perception of eye strain and pressure felt when marking on the screen compared with PBM, the feedback and support obtained from the OSM system, and overall satisfaction with OSM.

The 2013 study investigated the psychometric properties of the two scales using Rasch measurement analysis, which established the questionnaire as a robust instrument with Rasch person reliabilities of 0.80 for both scales (Yan & Coniam, 2013). The 2014 and 2016 studies then investigated whether markers' characteristics influenced their attitudes towards OSM (Coniam & Yan, 2016; Yan & Coniam, 2014). Results from the 2014 and 2016 studies showed markers' generally positive attitudes. The 2014 study reported that markers marking in both English and Chinese were more positive about OSM than those marking only in English or in Chinese; males were more positive than females; and older markers showed higher levels of acceptance of OSM than younger markers though age did not affect perceived ease of use in the OSM environment.

A somewhat vexed issue was identified from the 2016 study (Coniam & Yan, 2016) that markers in social science subjects (e.g., BAFS, HMSC, geography, biology, and history) reported less positive attitudes towards OSM than those in science-technology subjects (e.g., ICT, mathematics, and physics). Subject area was related to how scripts were divided up for marking, which in turn was associated with markers' attitudes. In science-technology subjects, most of the questions were closed ended with standardised responses. Markers in these subjects generally marked scripts by section (a set of questions with standardised responses); these markers reported more positive attitudes compared with markers in social science subjects. In the latter group of subjects, scripts typically consisted of a mixture of extended response (paragraph or essay questions), structured response (writing one or two lines), and standardised response (single-word, multiple-choice) questions. Markers in social science subjects typically were given single questions to mark instead of whole sections. It was unclear from the 2016 study findings whether these markers showed less positive attitudes towards OSM because of the boredom of marking just a single question, or because of the higher cognitive demand when marking open-ended questions with variable answers, since these factors were not included in the scales about markers' attitudes.

As such, the previous studies touched on some of the influential factors of reliability (e.g., usability of technology and feedback for markers) (Black et al., 2011; see Figure 1) in an OSM context. Nonetheless, a fine-tuned picture of how these factors contribute to – or reduce – reliability in OSM has yet to be built, which constitutes a major purpose of the current paper.

Research questions

In previous studies, the important question arose of what might be the reasons behind markers' less positive attitudes towards OSM in the subjects of BAFS, HMSC, geography, biology, and history. In this light, the current study sets out to explore:

RQ1. Why did markers in the five selected subject areas show less positive attitudes towards OSM than markers in other subjects?
RQ2. How were markers' attitudes related to the influential factors of reliability in OSM?
RQ3. How may OSM processes be improved for the enhancement of marking reliability?

Methods

To explore markers' perceptions about OSM in the HKDSE, the researchers conducted semi-structured telephone interviews with 31 markers from the subjects of BAFS, HMSC, geography, biology, and history. They were all secondary school teachers who responded to the survey that investigated markers' attitudes towards OSM as reported in Coniam and Yan's (2016) study.

Participants

The 31 participants (16 males and 15 females) were invited from 48 respondents to the survey (Coniam & Yan, 2016) who expressed willingness to join follow-up interviews;

this gave a response rate of 65% (non-responses were due to the teachers changing their mind or not answering the telephone calls). Participants were purposefully sampled from the five subject areas in which markers showed the least positive attitude towards OSM compared with markers from 14 other subjects (Coniam & Yan, 2016).

Participants' background information was extracted from survey responses (see Appendix 1). Each participant was assigned a pseudo name according to his/her gender and subject area, taking the first letter of the pseudo name as mnemonic. For example, **G**eoff was a male teacher of **g**eography; **B**art was a male teacher of **B**AFS, and **H**eidi was a female teacher of **H**MSC. To avoid confusion about participants' subject backgrounds, the initial letter for biology teachers' pseudo names was "c" instead of "b" (e.g., **C**lara was a female biology teacher), and the history teacher was named as **D**ianna to distinguish that teacher from HMSC teachers. As shown in Appendix 2, half of the participants were from BAFS ($n = 16$), while the rest were from the other four subject areas. The majority ($n = 22$) had participated in the HKDSE OSM exercise twice or more, while the rest ($n = 9$) had been OSM markers only once.

Data collection

The interviews tapped into participants' perceptions of their OSM experience via telephone interviews in Cantonese (the Chinese dialect spoken in Hong Kong) guided by a semi-structured protocol (see Appendix 3). Despite the absence of visual cues such as participants' facial expressions, body language, attire, and the surrounding environment that help with interpretation of their meanings (Ward, Gott, & Hoare, 2015), telephone interviewing offers the advantages of making participants feel at ease by participants being in a familiar environment (Novick, 2008).

The interview questions asked participants about their overall rating and satisfaction concerning the OSM experience, their perceptions of marking open- or closed-ended questions, their preference between OSM and PBM (paper-based marking), the perceived advantages and disadvantages of the quality control mechanisms and OSM technology, and their suggestions for improvements.

Prior to their telephone interviews, the participants completed written consent forms, which were mailed to them and then mailed back to the researchers by post. Participants were also provided with an information sheet explaining the objectives of the study, how they would be involved in the interviews, whether there would be potential risks for them, how the data and their personal information would be handled to protect confidentiality, and their rights to withdraw from the study without any adverse consequences. All interviews were audio-recorded upon seeking participants' verbal approval.

Data analysis

Transcription of the interviews was conducted word for word in Chinese, after which a bilingual member of the research team translated the completed transcripts from Chinese into English. Another bilingual member then reviewed the translated transcripts to verify the accuracy of translation.

The transcripts were analysed in two ways. First, simple numeric counting of participants' responses was conducted to provide an overall picture of their experiences and

attitudes regarding OSM, as presented in Appendix 2. Participants' experiences are further analysed in the Findings and discussion section.

Second, the interview data were analysed via thematic analysis in three stages for in-depth findings (Braun & Clarke, 2006). In Stage 1, one of the researchers conducted initial analysis by summarising the main points of all participants' responses and condensing the main points into categories and sub-categories. The categories and sub-categories were meaningful clusters; for example, under the category "Reasons for preferring marking on paper" were a number of sub-categories (e.g., "the quality assurance issue of giving markers reflection time when marking open-ended questions"). The categories and sub-categories served as codes for coding the transcripts. Stage 1 included four sub-steps: (1) analysing a set of six transcripts to derive an initial set of codes – these codes were reviewed by the other two researchers against the six transcripts to suggest changes to the codes when necessary; (2) coding the rest of the transcripts with the initial set of codes; (3) adding, combining, or removing the initial codes, so that the meanings expressed by participants were fully reflected by the codes; (4) creating a summary table of all codes that were illustrated by typical quotes drawn from the transcripts. In Stage 2, the researchers discussed the summary table with the codes and selected quotes and referred to the original transcripts when resolving disagreements on the coding results. In Stage 3, the researchers synthesised the codes representing participants' meanings into main themes, which will next be presented.

Findings and discussion

This section discusses three major themes arising from the interviews with markers in the five subject areas. The framework on the interaction of influential factors of reliability in OSM provides a conceptual background for interpreting the themes (adapted from Black et al., 2011; see Figure 1).

Theme 1: participants' overall experience and attitudes regarding OSM

Analysis of the summary of participants' experiences and attitudes towards OSM (see Appendix 2) provides several observations:

a. Overall rating of OSM experience

Echoing previous findings on HKDSE markers (Yan & Coniam, 2013, 2014), most participants showed a generally positive attitude by rating their OSM experience – on a 10-point scale between 1 (*very low*) and 10 (*very high*) – at a higher score of 7 or 8 ($n = 20$) or a neutral score of 5 or 6 ($n = 8$). Just a couple of participants however rated their experience at a relatively lower score of 4 ($n = 2$). Compared with the geography teachers who uniformly gave higher scores regarding their OSM experience, rating by participants in other subjects was fairly evenly spread across the high and middle ranges, with just three BAFS teachers and one history teacher giving neutral scores. The attitude ratings are presented in Appendix 2, under the sub-heading "a. Overall rating of OSM on a scale of 1–10: 1 = *Very Low*; 10 = *Very High*".

b. Overall satisfaction with OSM quality control mechanisms

A similar pattern to participants' rating of the OSM experience was identified regarding satisfaction with quality control mechanisms[1] (see Lee, 2016, for an overview of quality control mechanisms of OSM system for the HKDSE adopted by HKEAA). The majority ($n = 22$) expressed satisfaction; several ($n = 6$) were neutral; and just three reported dissatisfaction. Based on the participants' rating of OSM and their levels of satisfaction with quality control mechanisms, it can be inferred that the majority of markers accepted OSM as an assessment process that was capable of assisting them to attain reliability.

c. Marking by question versus marking by section

Two thirds of participants ($n = 19$) reported marking scripts *by question* (an open-ended question), including all 16 BAFS teachers, two geography teachers, and the only history teacher. The other 12 participants marked *by section* (a set of closed-ended questions), including all five biology teachers, all three HMSC teachers, and four geography teachers.[2] On the issue of marking by question, participants were divided when rating their OSM experience, whereas participants marking by section consistently gave high/medium scores. This is understandable, since marking open-ended questions requires comprehension of candidates' varyingly constructed responses based on the mark scheme, which places higher cognitive demand on markers than marking closed-ended questions to which candidates' responses are standardised (Massey & Raikes, 2006). The above findings of Section (c) are presented in Appendix 2, under the sub-heading "c. The way of marking experienced in OSM: marking by question or by section".

The pattern in participants' preferences for the method of marking showed quite a spread. About half ($n = 15$) indicated preferring marking by question due to the benefits of (a) enhanced accuracy, (b) better concentration, and (c) faster marking. A third ($n = 10$) preferred marking by section because this mode of marking: (a) helped them to have a better understanding of candidates' performance in the cohort, (b) helped them to become quickly familiarised with examination questions, and (c) made marking less boring. Several participants ($n = 6$) showed no preference. Given this result, the majority of participants did not see marking by question to be particularly disadvantageous. This result was possibly not a surprise since participants were mainly teachers of social science subjects, where open-ended questions are commonly used in assessments (Meadows & Billington, 2005). It should be noted that the participants' comments on this aspect of OSM were based on their theoretical positions, since most of them only experienced one of the two marking modes.

d. The mode of marking that participants preferred

Among the 25 participants who had experience of marking paper-based scripts, slightly less than half of them ($n = 11$) preferred OSM; another half ($n = 13$) preferred PBM; and a biology teacher showed no preference. Since the participants preferring PBM slightly outnumbered those preferring OSM, this helps to explain the previous survey result that markers in the five selected subjects were less positive towards OSM than those in other subjects (Coniam & Yan, 2016).

Based on Theme 1, markers' rating of OSM experience and overall satisfaction might be potentially related to influential factors of reliability at *the question and mark scheme level* (see Figure 1). It should be noted that the participants' subjective views should not be taken as evidence that such a relationship existed in the context of OSM in HKDSE. Further, markers marking closed-ended questions were more positive towards OSM compared with those marking open-ended questions. It might be possible that markers' varying attitudes towards OSM might be associated with their differing perceptions of marking reliability for open- versus closed-ended questions. Nonetheless, it was also likely that closed-ended questions can be more efficiently marked by OSM; and therefore the markers marking these questions held more positive attitudes towards OSM. The above-summarised findings thus warrant further research in the context of OSM through both quantitative and qualitative studies. Furthermore, some participants' preference for PBM implies their resistance to OSM – their reasons are explored in more detail under Theme 3.

Theme 2: perceived advantages of OSM compared with PBM

Participants commented on the advantages of OSM in relation to reliability from two aspects, namely, (a) quality control mechanisms and (b) usability of OSM technology.

a. Quality control mechanisms

Echoing researchers' expectation on OSM's potential to increase reliability (Bejar, 2011; Meadows & Billington, 2005), for many participants, the quality control processes of OSM afforded two reliability-related advantages:

- improved reliability and fairness for candidates;
- increased feasibility of double marking.

The majority of participants ($n = 22$) believed that reliability in OSM was enhanced because of the implementation of quality control mechanisms (see Note 1), which comprised training and standardisation through which marking was constantly monitored by the Chief or Assistant Examiners.

> The scripts are sampled for checking; once in a while qualifying (control) scripts are sent to you. If you can't pass the qualifying round, you will be asked to do re-training by marking another set of control scripts. The mechanism should be continued. (Bart)

Quality control mechanisms were also believed to enhance the feasibility of double-marking in OSM:

> I used to mark scripts that were handled manually (on paper). You could conduct double marking but it was relatively difficult and only a small number (of scripts) could be double marked. Perhaps you couldn't get instant feedback. It is good that computers are used to allow double marking in OSM. (Gerald)

Three participants mentioned that since OSM increased marking accuracy compared with PBM, this in turn improved fairness of assessment for candidates.

> OSM is more accurate as you make fewer mistakes with marking. Thus OSM is fairer to candidates. (Bambi)

b. Usability of OSM technology

Some participants mentioned that OSM technology increased their work efficiency by offering three advantages as stated below, which helped them to maintain accurate and consistent marking:

- better concentration;
- real-time access to marking statistics;
- automated OSM processes, including:
  - sorting and loading scripts, and having the marking rubric displayed on the screen at all times;
  - being able to adjust font sizes of the scripts on the screen;
  - calculating and recording of marks.

About a third of participants ($n = 11$) commented that the OSM centre environment was a factor that contributed to concentration during marking, which in turn enhanced marking accuracy and speed. Such comments correspond to previous research evidence that improved concentration can increase reliability by removing distractions for markers (Tisi et al., 2013).

> Marking with a computer makes you concentrate better. The environment (in OSM centres) makes you concentrate better. (Gerry)

> Taking scripts home to mark increased the risk of losing scripts and reduced concentration. Thus I prefer OSM more. (Heidi)

Three participants mentioned the convenience of having real-time access to the statistics about one's own marking and about other markers' marking. They felt that by viewing the statistical information, they were able to self-monitor their marking performance, which assisted in consistent marking.

> The advantage is that you can compare statistics and maintain consistency. (Hettie)

> Marking (on screen) is faster, and you can view statistical information, such as how many minutes were spent on marking a script, and the grade of each script. (Barclay)

While the above observation by participants is reassuring, it would be necessary to further investigate the process in which markers self-improve marking performance through self-monitoring in future research.

Four BAFS teachers noted that the automated OSM processes improved their work efficiency; such an observation concurs with the notion that automated processes exemplify technological advances in assessments, which help develop reliability (Ramakrishna, Sree, Harish, Swarna, & Vasundhara, 2012).

> The advantage is that marks are recorded automatically, which makes administrative matters more convenient. (Barry)

> You don't have to look for information, because the mark scheme is automatically loaded to the scripts being marked. (Basia)

In short, for the above-quoted participants, technological advancements made OSM superior to PBM in making it possible to improve marking reliability and work efficiency

(Ramakrishna et al., 2012; Zhang et al., 2003). Detailed analysis of whether and how markers' self-monitoring and self-improving of their marking can enhance inter-rater and intra-rater reliability in the OSM system appears to be a promising future research direction.

To summarise Theme 2: Participants mostly associated OSM's advantages to its technology-enhanced mechanisms, procedures, and features. The findings thus provide detailed information of how the participants perceived the advantages of OSM with regards to marking reliability.

Theme 3: perceived disadvantages of OSM compared with PBM

In relation to marking reliability, participants were asked to comment on disadvantages of OSM regarding: (a) quality control mechanisms and (b) usability of OSM technology.

a. Quality control mechanisms

Six participants expressed reservations of three kinds:

- issues related to OSM quality control mechanisms:
  - prescribed answers to questions in control scripts being occasionally incorrect or causing controversy;
  - the monitoring procedure being potentially over-sensitive to minor marking errors when scripts were divided into many portions;
- a lack of opportunity for discussion with other markers when marking open-ended questions;
- a lack of time for reflection when marking responses to open-ended questions.

One participant cast doubt on the prescribed correct answers to certain questions in the control scripts and suggested careful selection of control scripts to avoid potential disputes during marker training. This observation echoes researchers' advice that control scripts (i.e., exemplars) used for training and standardisation should represent the full range of candidate responses in order to inform markers' judgments (Meadows & Billington, 2005).

> Sometimes when I marked control scripts, the prescribed answers were different from the correct answers, but I couldn't get qualified. The technicians were unable to fix the problem even when the Assistant Examiner said I was right … The scripts for qualifying should not contain questions that might cause controversies. (Bart)

For three other participants, the monitoring procedure was overly stringent, since it was sensitive to tiny discrepancies in marking, which was attributed to the separation of an examination paper into many small portions. They were worried that this might cause the OSM system to unduly penalise tiny marking discrepancies. HKEAA might need to investigate this issue of quality control and take precautions in order to avoid discouraging markers from making legitimate judgment based on their expertise (Newton, 2013).

> Markers get investigated even when there are only one or two marks' differences. The quality control mechanism gets a bit abused. (Gerry)

> The scripts in my subject are divided into very small portions. There is a greater chance for us to give wrong marks. Because when you are marking a small question, even with one or two marks' difference there may be quite a big discrepancy. (Baxter)

Another participant mentioned the need for the opportunity to discuss assessment-related issues among markers, such as marking standards and practices, especially when marking open-ended questions. This echoes previous research evidence that reliability may be maintained by engaging markers in a community of assessment practice, which allows them to share a common set of knowledge and practice, such as how the mark scheme is interpreted (Baird et al., 2004).

> Markers should be allowed to participate in discussions about questions and the marking. We need to understand marking standards and the overall trend of colleagues' marking … and how the curriculum determines the assessment. This is where I would like to see improvement. (Barbara)

Two teachers in BAFS commented that marking the highly variable responses to open-ended questions was more accurate when conducted on paper rather than on screen, because such responses required more time for digestion and reflection.

> When marking on paper, you can read the answers several times and revise marks. When marking on screen, I have to handle a larger number of scripts and it is more tiring. I feel the accuracy of OSM may be lower than marking on paper. (Basia)

While the above-quoted marker perception concurs with previous evidence that markers applied sophisticated cognitive strategies in marking open-ended questions, it should also be noted that markers were found in previous research to use different cognitive strategies for OSM than strategies for PBM in order to sustain reliability (Suto et al., 2011).

Another biology teacher suggested that compared with PBM, it was harder to retrieve scripts for further discussion with the coordinator, which might reduce marking accuracy.

> It is more accurate and efficient to mark on paper. In the past, I could make a note on a script, but I can't do that now. It is difficult to retrieve a script if I want to discuss with the coordinator some aspects about the script. (Collin)

The above-quoted comments resonate with previous findings that marking in OSM poses considerable cognitive demand due to the need to read from the screen (Campbell, 2005), even though the reliability of OSM is comparable with PBM regardless of examination questions being open or closed ended (Johnson, Nádas, & Bell, 2010). These marker concerns over potential risks in the quality control process reinforce the need to facilitate marker agreement not only through careful selection of control scripts and sufficient feedback from Chief and Assistant Examiners, but also through markers' peer discussions for collective assessment knowledge building.

b. Usability of OSM technology

Participants alluded to four disadvantages of OSM in this regard:

- eye tiredness as a result of facing the computer for a long time;
- blurred scripts due to the scanning of scripts or using black-and-white fonts in scripts;
- problems with the OSM system software;
- problems with workstation hardware.

Nine out of the 31 participants mentioned that eye tiredness occurred after marking on the screen for a long time. A participant warned that this disadvantage might risk the HKEAA losing experienced examiners.

> Since OSM was implemented, markers have been relatively young, as it is difficult for older teachers. In contrast, when marking on paper, markers were mostly experienced teachers. The loss of experienced markers might become even worse in the future. (Bambi)

Although the above-quoted remark contradicted previous survey result that older markers reported greater ease of use in the OSM environment than younger markers (Coniam & Yan, 2016), eye tiredness would indeed appear to be a persistent problem.

Three other participants mentioned the problem of unclear images of scanned scripts and suggested using coloured fonts in examination papers and improving computer screen resolution. Whereas this was a technical issue, existing research suggests the issue is likely to induce risks of lowered reliability, which should be closely monitored in public examinations (Black, 2010).

> Some scripts were not clear enough after scanning; sometimes the use of correction pens blurred the scripts. Possible solutions are to explore if the papers can possibly be printed in colour or if the resolution can be clearer. (Bailey)

Nearly a third of participants ($n = 9$) reported various problems with the OSM software and hardware. Problems regarding the software mainly concerned difficulties in using certain annotation symbols (such as the tick symbol for correct answers) ($n = 4$), while one participant mentioned needing more symbols for annotation. This is understandable, since annotation symbols play an important role in markers' comprehension of candidates' responses and communication of judgments to other markers (Campbell, 2005; Johnson, Hopkin, Shiell, & Bell, 2012). Thus, examination organisations need to address such concerns by improving the usability of the annotation software.

> Using the mouse to change between tick and cross was not very convenient. (Hettie)

> There needs to be more annotation symbols to choose from. Sometimes what markers want to express can't be expressed by the symbols currently available. (Bates)

> Symbols for marking should be adhered to closely, because the HKEAA has its requirements … There is a symbol called "irrelevant" for annotating repeated answers. I remember once I didn't put the symbol against such answers – and I was asked to re-mark the script. (Gerald)

Regarding the hardware, participants mentioned problems such as computers breaking down during marking ($n = 3$), and needing mouse pads to avoid wrist strain ($n = 1$), which reduced their work efficiency.

> A few times the computer could not load the scripts until I rebooted it. The marking settings would then be lost. Resetting is pretty inconvenient. (Geoff)

Based on the above-quoted comments, it appears that improving the monitor resolution to make marking on screen more comfortable for markers, increasing the clarity of scanned scripts, and improving the annotation symbols are three key aspects of improving the usability of OSM technology in order to enhance marking reliability.

To sum up Theme 3: The markers identified certain aspects of the quality control mechanisms and OSM technology that should be improved, which coincides with researchers'

recommendations for ensuring reliability in OSM. In particular, markers' concerns over the quality control mechanisms point to the need for examination organisations to maximise their support for markers in attaining greater reliability and building assessment expertise.

Conclusion

The current study has set out to investigate specific issues identified from markers' OSM experience in relation to key influential factors of marking reliability in OSM. To suit the context of OSM and fulfil the purpose of the current study, the Black et al. (2011) model of influential factors of marking reliability has been adapted by incorporating specific factors related to marking reliability in OSM that were identified in three previous studies (Coniam & Yan, 2016; Yan & Coniam, 2013, 2014) (see the adapted framework in Figure 1). An important conceptual contribution of the current paper is then the identification and detailed examination of the influential factors that can be managed by the examination authority (Type 1 factors in the adapted framework; see Figure 1) with a view to enhancing marking reliability in OSM.

Emerging from the qualitative data were three major themes pertaining to markers' overall experience and satisfaction, and the advantages and disadvantages they perceive about the OSM system in the context of HKDSE.

Theme 1 shows that there might be a potential association between markers' attitudes towards OSM (i.e., rating of and satisfaction with OSM experience) and the types of questions they marked. Markers who dealt with closed-ended questions were positive towards OSM, while those marking open-ended questions were rather divided in their attitudes. Previous research indicates higher reliability for closed-ended questions than open-ended questions – the latter being more cognitively demanding by requiring marker interpretation of the mark scheme (Black, 2010). In this sense, Theme 1 implies the likelihood of markers' attitudes being potentially associated with influential factors of reliability at the question and mark scheme levels (see Figure 1). It should be cautioned that the qualitative findings based on the interview data would not suffice to empirically establish the interconnection between marker attitudes and marking reliability in OSM. In the light of this limitation of the current study, one important direction for future research is to investigate more explicitly whether or not markers' attitudes are indeed interrelated with marking reliability in OSM context. The qualitative finding raises a practical issue that, since most participants were from social science subjects, managing markers' expectations of their OSM experience in these subjects may require attention by examination organisations.

Theme 2 concerned the perceived advantages of OSM with regards to quality control and usability of technology. Concurring with previous research (cf. Coniam & Yan, 2016), the majority of markers appreciated the potential of OSM for attaining greater reliability as compared with PBM. This finding, however, was based on markers' subjective perceptions on marking reliability in OSM and PBM; thus, it does not constitute empirical evidence of comparison between the levels of reliability in OSM and PBM and should be interpreted with caution. According to interviews of a few markers, one advantage afforded by the OSM technology was markers' self-monitoring of marking consistency using real-time marking statistics. Acknowledging that the interview protocol of this study could have been expanded to investigate markers' experiences of self-monitoring

in the OSM system, such marker behaviours might profitably be further investigated to inform strategies for promoting markers' self-regulative actions in attaining reliability. Together, the findings under Theme 2 showed that OSM potentially possesses distinctive advantages that are related to quality control processes and usability of OSM technology, which are two important influential factors in marking reliability as demonstrated in the adapted framework following Black et al. (2011; see Figure 1, in which these two factors are categorised under the "Organisation of the OSM task").

Finally, under Theme 3 some markers identified various disadvantages associated with the quality control process and usability of technology in OSM. The disadvantages ranged from the use of disputable control scripts for training and standardisation, through insufficient time for reflecting on open-ended questions and insufficient opportunities for discussion with other markers, to defects of OSM annotation symbols and unclearly scanned scripts. Such disadvantages were perceived by markers to be possible sources of unreliability, which coincide with researchers' caution against threats to reliability in assessment generally (Tisi et al., 2013). As such, there is a need to further examine the influential factors at the level of the quality control and usability of OSM technology (see Figure 1) as a way to minimise risks to marking reliability in OSM.

Potential strategies for addressing the above-mentioned disadvantages may be implemented by examination organisations with a view to enhancing reliability in OSM. These include (a) fine-tuning the quality control mechanisms to avoid penalising minor marking errors; (b) consulting markers on enhancements to the software and hardware of OSM workstations; and (c) creating face-to-face or virtual communities of assessment practice via means such as forums, workshops, and marker meetings to facilitate markers' collective assessment knowledge building.

Although this study did not touch on the role of the factors involved in the logistical arrangements and the OSM environment in attaining marking reliability[3], such factors remain crucial for safeguarding the security of candidates' scripts and a comfortable environment for markers (Adams, 2005). Thus, these factors should be considered as an integral part of the quality assurance plan in OSM.

Whether marking on screen or on paper, the human marker will continue to play a significant part in the overall picture of marking reliability because of the need to assess a wide range of candidate performance and understanding in order to address important curricular objectives, such as the development of critical thinking and problem solving (Ahmed & Pollitt, 2011). The significance of the current study lies in its further development of an existing framework suggested by Black and colleagues (2011) on influential factors of marking reliability in OSM. Although the original framework has been proposed by Black and colleagues (2011) as a causal model, this study has shown the utility of the revised framework for deriving qualitative findings that contribute to new understanding of the relationship between markers' attitudes towards OSM and the reliability of marking and how associated factors might interact to affect reliability. As mentioned earlier, the present study is qualitative and uses perception-based data from knowledgeable participants, that is, who were OSM markers. Such qualitative research has its own canons of reliability, and serve as a valuable complement to the work of Black et al. (2011). Finally, it is acknowledged that the exploratory nature of findings derived from markers' interviews limits the capacity of the current study to empirically examine the potential linkages between the influential factors and marking reliability in OSM. Nonetheless, the qualitative

findings usefully point towards several avenues for further investigation as explicated in the preceding discussion, which in turn will assist in advancing the theorisation of influential factors of marking reliability in OSM.

Notes

1. OSM quality control mechanisms used by the HKEAA for the HKDSE include: (a) A *qualifying procedure*, which uses standardisation scripts to determine a marker's marking accuracy against prescribed answers and whether the marker may therefore qualify for marking. Qualifying involves: the *training stage*, where markers explore how the OSM system operates (e.g., viewing reference scripts that have been assigned "gold standard" grades); and the *qualifying stage*, where markers mark qualifying scripts with prescribed grades. (b) A *double marking* procedure, which is applied with subject panels where test papers include open-ended questions/items. Double marking involves each script being marked by two markers. If the discrepancy between the two markers' marks exceeds a prescribed limit, a third (and sometimes a fourth) marker will be invoked.
2. The difference in marking reported by the geography teachers was because they were referring to OSM marking in different years. The method of marking in geography changed from marking by section to marking by question.
3. In their interviews, participants were asked to share their opinions about the logistical arrangements and the OSM environment. Nonetheless, participants did not report the two aspects as directly relevant to marking reliability. Therefore, this part of the data is not reported in the current paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Min Yang is an assistant professor in the Department of Curriculum and Instruction at The Education University of Hong Kong. She has researched into technology-enhanced feedback and assessment approaches to enhancing student learning.

Zi Yan is an assistant professor in the Department of Curriculum and Instruction at The Education University of Hong Kong. His main research interests are in educational and psychological assessment, especially the application of Rasch analysis.

David Coniam is a chair professor in the Department of Curriculum and Instruction at The Education University of Hong Kong, where he is a teacher educator, working with teachers in Hong Kong secondary schools. His main publication and research interests are in language assessment, language teaching methodology, and corpus linguistics.

ORCID

Zi Yan http://orcid.org/0000-0001-9305-884X

References

Adams, C. (2005, September). *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", Abuja, Nigeria.

Ahmed, A., & Pollitt, A. (2011). Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy & Practice*, *18*(3), 259–278. doi:10.1080/0969594X.2010.546775

Baird, J.-A., Greatorex, J., & Bell, J. F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 331–348. doi:10.1080/0969594042000304627

Bejar, I. I. (2011). A validity-based approach to quality control and assurance of automated scoring. *Assessment in Education: Principles, Policy & Practice*, *18*(3), 319–341. doi:10.1080/0969594X.2011.555329

Black, B. (2010, August). *Investigating seeding items used for monitoring on-line marking: Factors affecting marker agreement with the gold standard marks*. Paper presented at the International Association for Educational Assessment 36th Annual Conference, Bangkok, Thailand.

Black, B., Suto, I., & Bramley, T. (2011). The interrelations of features of questions, mark schemes and examinee responses and their impact upon marker agreement. *Assessment in Education: Principles, Policy & Practice*, *18*(3), 295–318. doi:10.1080/0969594X.2011.555328

Bramley, T. (2007). Quantifying marker agreement: Terminology, statistics and issues. *Research Matters: A Cambridge Assessment Publication*, *4*, 22–28.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. doi:10.1191/1478088706qp063oa

Campbell, A. (2005). Application of ICT and rubrics to the assessment process where professional judgement is involved: The features of an e-marking tool. *Assessment & Evaluation in Higher Education*, *30*(5), 529–537. doi:10.1080/02602930500187055

Coniam, D., & Yan, Z. (2016). A comparative picture of the ease of use and acceptance of onscreen marking by markers across subject areas. *British Journal of Educational Technology*, *47*(6), 1151–1167. doi:10.1111/bjet.12294

Haggie, D. (2008). *The strategic use of marking technologies to support innovation and diversity in assessment*. Singapore: RM Education plc.

Johnson, M., Hopkin, R., Shiell, H., & Bell, J. F. (2012). Extended essay marking on screen: Is examiner marking accuracy influenced by marking mode? *Educational Research and Evaluation*, *18*(2), 107–124. doi:10.1080/13803611.2012.659932

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. doi:10.1111/j.1467-8535.2009.00979.x

Lee, C. (2016). The role of the Hong Kong Examinations and Assessment Authority. In D. Coniam & P. Falvey (Eds.), *Validating technological innovation: The introduction and implementation of onscreen marking in Hong Kong* (pp. 9–21). Singapore: Springer.

Massey, A. J., & Raikes, N. (2006, September). *Item-level examiner agreement*. Paper presented at the Annual Conference of the British Educational Research Association, University of Warwick, Coventry, UK.

Meadows, M., & Billington, L. (2005). *A review of the literature on marking reliability*. Unpublished AQA report produced for the National Assessment Agency.

Newton, P. E. (2013). Ofqual's Reliability Programme: A case study exploring the potential to improve public understanding and confidence. *Oxford Review of Education*, *39*(1), 93–113. doi:10.1080/03054985.2012.760285

Novick, G. (2008). Is there a bias against telephone interviews in qualitative research? *Research in Nursing & Health*, *31*(4), 391–398. doi:10.1002/nur.20259

Ramakrishna, A., Sree, B. N., Harish, P. S., Swarna, S., & Vasundhara, C. H. (2012). Design and implementation procedure for administration and evaluation in e-marking system. *International Journal of Innovative Technology and Exploring Engineering*, *2*(1), 15–19.

Suto, I., Nadas, R., & Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education*, *26*(1), 21–51. doi:10.1080/02671520902721837

Tisi, J., Whitehouse, G., Maughan, S., & Burdett, N. (2013). *A review of literature on marking reliability research*. Slough: NFER.

Ward, K., Gott, M., & Hoare, K. (2015). Participants' views of telephone interviews within a grounded theory study. *Journal of Advanced Nursing*, *71*(12), 2775–2785. doi:10.1111/jan.12748

Yan, Z., & Coniam, D. (2013). Assessing the ease of use in the environment and markers' acceptance of on screen marking: A Rasch measurement perspective. *Educational Research & Evaluation*, *19*(5), 461–483. doi:10.1080/13803611.2013.793604

Yan, Z., & Coniam, D. (2014). The effects of key demographic variables on markers' perceived ease of use and acceptance of onscreen marking. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 464–480. doi:10.1080/0969594X.2014.953910

Zhang, L Y., Powers, D. E., Wright, W., & Morgan, R. (2003). *Applying the online scoring network (OSN) to advanced placement program (AP) tests* (ETS Research Report No. RR-03-12). Princeton, NJ: Educational Testing Service.

Appendix 1. Participants' background information ($n = 31$)

Name	Gender M = male F = female	Subject	Experience with OSM in HKDSE		Experience with PBM	Name	Gender M = male F = female	Subject	Experience of OSM in HKDSE		Experience with PBM
			Once	Twice or more					Once	Twice or more	
Bahar	F	BAFS	X	X	X	Clara	F	biology	X		X
Barley	F	BAFS	X	X	X	Catherine	F	biology	X		X
Bambi	F	BAFS	X	X	X	Calvin	M	biology	X		X
Baptista	F	BAFS	X	X		Christopher	M	biology	X	X	
Bara	F	BAFS	X	X	X	Collin	M	biology	X		X
Barbara	F	BAFS	X		X	Dianna	F	history	X		X
Basia	F	BAFS	X	X	X	Gemma	F	geography	X	X	X
Bathsheba	F	BAFS	X	X	X	Georgina	F	geography	X	X	
Bailey	M	BAFS	X	X	X	George	M	geography	X	X	
Barry	M	BAFS	X	X	X	Geoff	M	geography	X	X	
Bart	M	BAFS	X	X	X	Gerald	M	geography	X		X
Bates	M	BAFS	X		X	Gerry	M	geography	X	X	X
Baxter	M	BAFS	X	X	X	Heidi	F	HMSC	X	X	X
Barrington	M	BAFS	X	X	X	Helena	F	HMSC	X	X	
Barlow	M	BAFS	X		X	Hettie	F	HMSC	X	X	X
Barclay	M	BAFS	X	X	X						

Notes: All participants' names are disguised using pseudo names. "OSM" and "PBM" stand for on-screen marking and paper-based marking, respectively. "HKDSE" refers to the Hong Kong Diploma of Secondary Education. "BAFS" is the short form for business, accounting, and financial studies. "HMSC" refers to health management and social care.

Appendix 2. Summary of participants' experiences/attitudes regarding OSM

I. Background of participants (reported in Methods section)	
a. Total number of participants and their subject areas	• 31 teachers, including: *16 BAFS teachers, 6 geography teachers, 5 biology teachers, 3 HMSC teachers, and 1 history teacher*
b. Involvement in OSM* of the HKDSE**	• 22 teachers participated in OSM twice or more times, including: *13 BAFS teachers, 5 geography teachers, 3 HMSC teachers, and 1 biology teacher* • 9 teachers participated in OSM only once, including: *3 BAFS teachers, 4 biology teachers, 1 geography teacher, and 1 history teacher*
II. Details of participants' responses related to Theme 1 (reported in Findings and discussion section)	
a. Overall rating of OSM on a scale of 1–10: 1 = *Very Low*; 10 = *Very High*	• Modal score = 7; Mean score = 6.67 • 20 teachers gave a high score of 7 or 8, including: *9 BAFS teachers, 6 geography teachers, 3 biology teachers, and 2 HMSC teachers* • 8 teachers gave a neutral score of 6 or 5, these included: *4 BAFS teachers, and 2 teachers of biology, history, and HMSC, respectively* • 2 teachers gave a negative score of 4, who were: *1 BAFS teacher and 1 history teacher*
b. Satisfaction with quality assurance mechanisms in the OSM system: whether participants were "satisfied", "dissatisfied", or "neither satisfied nor dissatisfied".	• 22 teachers expressed satisfaction, including: *10 BAFS teachers, 5 geography teachers, 4 biology teachers, 2 HMSC teachers, and 1 history teacher* • 6 teachers were neither satisfied nor dissatisfied, including: *4 BAFS teachers, 1 biology teacher, and 1 HMSC teacher* • 3 teachers showed dissatisfaction, including: *2 BAFS teachers and 1 geography teacher*
c. The way of marking experienced in OSM: marking by question or by section	• 19 teachers marked by question, including: *16 BAFS teachers, 2 geography teachers, and 1 history teacher* • 12 teachers marked by section, including: *5 biology teachers, 4 geography teachers, and 3 HMSC teachers*
d. Preferred way of marking: marking by question, or by section	• 15 teachers preferred marking by question, including: *10 BAFS teachers, 2 geography teachers, and 1 teacher of biology, history, and HMSC, respectively* • 10 teachers preferred marking by section, including: *6 BAFS teachers, 2 HMSC teachers, 1 biology teacher, and 1 geography teacher* • 6 teachers had no particular preferences, including: *2 biology teachers, 4 geography teachers*
III. Details of participants' responses related to Theme 2 and Theme 3 (reported in Findings and discussion section)	
a. Prior experience of PBM in public examinations	• 25 teachers had prior experience of PBM • 5 teachers did not mark by PBM, who were 3 geography teachers and 1 biology teacher, and 1 HMSC teacher; while 1 biology teacher's response was missing

b. Preferred mode of marking: either OSM or PBM	Among the 25 participants with prior PBM experience: • 11 teachers preferred OSM, including: *6 BAFS teachers, 2 geography teachers, 2 HMSC teachers, and 1 teacher of biology* • 13 teachers preferred PBM, including: *8 BAFS teachers, 3 biology teachers, 1 geography teacher, and 1 history teacher* • 1 teacher of biology showed no preference

Notes: "BAFS" means business, accounting and financial studies. "HMSC" means health management and social care. "HKDSE" refers to the Hong Kong Diploma of Secondary Education. "Way of marking" means dividing scripts by single questions or sections consisting of several questions for marking.

Appendix 3. Semi-structured interview protocol on teachers' perceptions regarding OSM

(1) How would you rate your experience of on-screen marking (OSM) on a scale from 1 (*very low*) to 10 (*very high*)?
(2) Did you conduct marking by question (marking one open-ended question), or by section (marking a set of close-ended questions)? Which of these two ways of marking they would you prefer, and why?
(3) Did you have prior experience of paper-based marking (PBM)? How would you compare OSM with PBM? Which of these two modes of marking would you prefer, and why?
(4) What are the advantages and disadvantages of OSM regarding the quality control mechanism? What suggestions would you give to improve such mechanisms? (Prompt: The quality control mechanisms include marker training and standardisation using control scripts, feedback provision by the Chief and Assistant Examiners, etc.)
(5) What are the advantages and disadvantages of the usability of OSM technology? What suggestions would you give to improve OSM technology? (Prompt: Examples of OSM technology include software and hardware of workstations, accessibility of marking information in the OSM system).