

A comparative picture of the ease of use and acceptance of onscreen marking by markers across subject areas

David Coniam and Zi Yan

David Coniam is a chair professor in the Department of Curriculum and Instruction at The Hong Kong Institute of Education, 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. Zi Yan is an assistant professor in the Department of Curriculum and Instruction at The Hong Kong Institute of Education. His main research interests are in educational and psychological assessment, especially the application of Rasch analysis. Address for correspondence: Prof. David Coniam, D1-1/F-38, Department of Curriculum and Instruction, The Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, N.T., Hong Kong. Email: coniam@ied.edu.hk

Abstract

Onscreen marking (OSM) has been used for the majority of Hong Kong public examinations since 2012. The current study compares marker reactions to OSM, ie, perceived ease of use and acceptance of OSM, against the backdrop of virtually all subject areas being marked on screen. The data were collected from three major sources: (1) survey data obtained from 1743 markers across 14 major subject areas, (2) markers' qualitative comments about the OSM system and (3) post-hoc interviews with a key informant from the Hong Kong Examinations and Assessment Authority (HKEAA). Results showed that, in general, markers revealed a high level of perceived ease of use and positive acceptance of OSM. The effect of subject area for both scales was statistically significant. On the *Ease of Use in the OSM Environment* scale, markers of information and communication technology (ICT) and mathematics were the most positive, with markers of history and geography the least positive. On the *Acceptance of OSM* scale, markers of mathematics and ICT were the most accepting, with markers of biology and geography the least. The analysis of survey data was triangulated by markers' qualitative comments together with the HKEAA staff interview. Possible explanations for the results are proposed and implications for the further development of OSM are briefly discussed.

Introduction

The rapid growth of information and communication technologies (ICTs) has had increasingly powerful influences on various aspects of education including assessment. A salient example is the shift from paper-based assessment to computer-based assessment as well as the shift from paper-based marking to onscreen marking (OSM) happening in higher education as well as high-stakes examinations.

In the UK, Cambridge Assessment (as it is now known) first experimented with the OSM of scanned paper scripts in 1999. After further piloting in 2006, Cambridge Assessment invested substantial sums into OSM for the 5-year period up to 2012 (Raikes, Greatorex & Shaw, 2004). RM Results (2014), an organisation specialising in onscreen testing and OSM, claims that 143 million exam pages were scanned and marked on screen in 2013 using RM Assessor, an OSM application. The official body of the UK Government *Ofqual* (the Office of Qualifications and Examinations Regulation) has declared that "onscreen marking is now the main type of marking

Practitioner Notes

What is already known about this topic

- From 2012, the majority of public examinations in Hong Kong have been marked on screen. Hong Kong markers had a positive perceived ease of use in the onscreen marking (OSM) environment and showed an increasing but slow acceptance of OSM (see Yan & Coniam, 2013).

What this paper adds

- The current study compares marker reactions to OSM, ie, perceived ease of use and acceptance of OSM, from the larger perspective of the subject area.

Implications for practice and/or policy

- The comprehensive picture of OSM presented in the current study provides a positive signal for the shift from paper-based marking to OSM in the Hong Kong context as well as for any jurisdiction where examinations are being marked on screen.
- The fact that certain subject areas are more or less positively disposed to OSM is an issue that the examination authorities will need to consider with regard to how OSM is presented in the future to markers of different subject areas.
- The results identify the major disadvantages of OSM as perceived by markers and provide concrete suggestions for future development of the OSM system.

for general qualifications in the UK,” with, in 2012, over 10 million examination scripts marked in such a manner in the UK.

In Africa, the global data capture specialist, the DRS Group (2014) reports a growing take-up of OSM in Africa, after the Zimbabwe School Examinations Council increased the speed of delivery and accuracy using DRS’s OSM technology. The Zimbabwe School Examinations Council launched its OSM scheme in June 2011, the first country in Africa to use OSM for public examinations. Eighty thousand O-level exam scripts were scanned for certain core subjects such as mathematics and integrated science in June 2011, rising to 600 000 scripts for additional subjects including accounting, physics and chemistry in November 2013. The success in Zimbabwe led to OSM also beginning to be adopted in other African countries such as Nigeria, Ethiopia and Tanzania.

It is in China, however—because of the sheer numbers of both candidates and markers—that the greatest use of OSM is made in the marking of public examinations (Luo & Liu, 2009; Ma, 2004). OSM was first introduced in the 1990s to mark subjective items in high-stakes tests with Guangxi being the first province to trial using OSM to mark its *gaokao* English test (the *gaokao* being China’s university entrance examination) in 1999 (Yang, 1999). This successful piloting of marking the *gaokao* via OSM strengthened the Chinese government’s confidence in adopting the technology for marking high-stakes tests nationwide. While exact figures are not readily available, the *gaokao* candidature—over 10 million in 2010—provides an overall perspective on OSM adoption. The success of the use of OSM in the marking of the *gaokao* spurred much wider use of OSM in China. In 2000, the Ministry of Education ordered that OSM be adopted to mark the *gaokao* in more provinces (Chang, 2000, June 13) and by 2006, 22 provinces and municipalities were employing OSM to mark the *gaokao* (Lu, 2006, June 26). Other high-stakes tests have also moved to using OSM, another notable example being the *zhongkao* (the senior high school entrance examination), which since 2010 has seen considerable uptake of OSM in many provinces (“First use of onscreen marking for Zhongkao in Qingdao—results release on 4 July”, 2014).

In Hong Kong, OSM is administered in a centralised approach, which is probably different from other systems. Markers for a given subject will be practising teachers in that subject. Teachers apply to mark and go through a rigorous selection and training process. Markers are paid at standard rates by the Hong Kong Examinations and Assessment Authority (HKEAA). The size of marker panels will vary according to the number of candidates in that subject. After candidates have completed an examination, the scripts are collected and delivered to the OSM centres. Here, the scripts are scanned and the images saved. Following this, markers go to an OSM centre of their choice (10 are currently operational across Hong Kong), where they mark at dedicated workstations (see Coniam, 2009 for more details). Images of answers are distributed around the system to markers for viewing and marking via a secure intranet. As well as marking on screen and entering marks at question or item level, markers are also able to annotate scripts without the second marker being affected by the first marker's comments, something they cannot do with paper-based marking. A screenshot (Figure 1) is provided to give readers a flavour of what the OSM system looks like. Basically, the marking interface includes three parts, ie, the scanned script, the marking board and the toolbar for marking. A brief explanation of the functions of the toolbar is also provided in the figure.

Accompanying the wide use of OSM is the increasing research attention on relevant areas including the comparability between paper-based marking and OSM (eg, Geranpayeh, 2011; Johnson, Hopkin, Shiell & Bell, 2012; Johnson, Nadas & Bell, 2010), the development of OSM

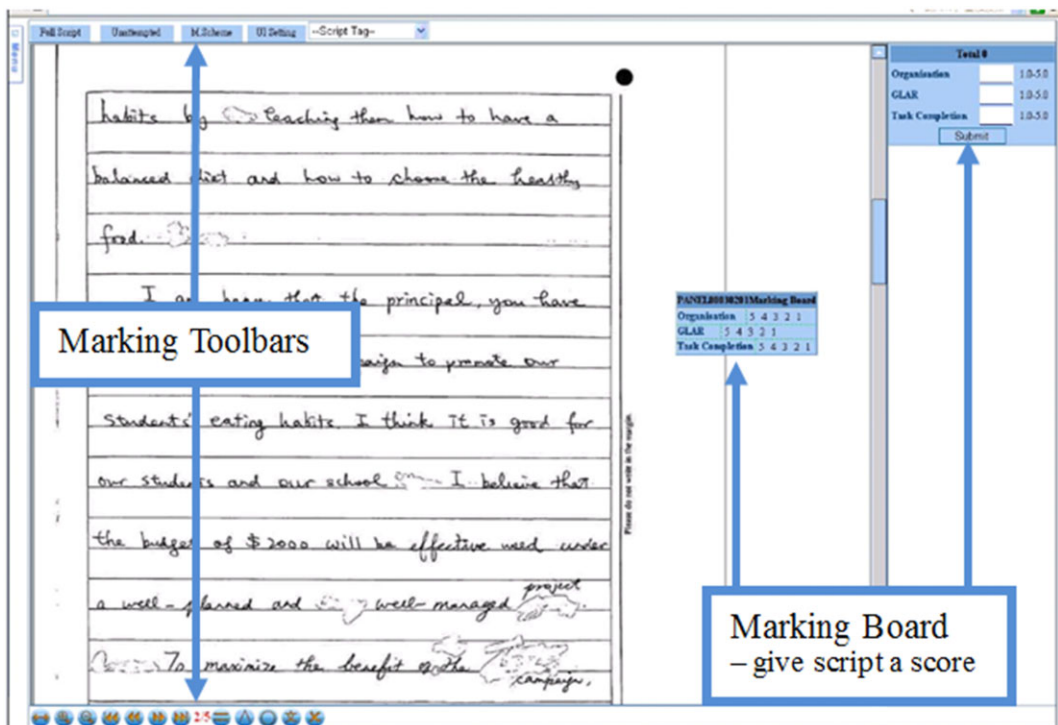


Figure 1: The marking interface and the functions of the toolbar. **Full Script**, view whole script; **Unattempted**, denote script as "Unattempted" (0 marks awarded automatically); **M.Scheme**, display Marking Scheme; **UI Setting**, define Marking Board personalized settings; **--Script Tag--**, add script tag (if desired) to script for future reference; , restore, enlarge or reduce size of script image; , navigate to first, previous, next or last page of script; **1,3**, first number indicates current page number; last number indicates total number of pages in the script; , add symbols and marking comments to script.

systems (eg, Campbell, 2005; Ramakrishna, Navya Sree, Sri Harish, Swarna & Vasundhara, 2012), as well as markers' attitudes towards OSM (eg, Coniam, 2011, 2013; Yan & Coniam, 2014). Nevertheless, the literature about marking issues in the new assessment environment remains limited.

The current study is the culmination of a number of previous studies (see, for example, Coniam, 2009, 2011, 2013; Coniam & Yeung, 2010; Yan & Coniam, 2013) investigating marker attitude towards OSM in the Hong Kong context where, from 2012 onwards, the majority of public examinations have been marked on screen. The unique feature of this study is that the study covers the data from all markers for all papers in all major subject areas marked on screen in the 2012 public examinations including both the Hong Kong Diploma in Secondary Education (HKDSE) and the Hong Kong Advanced Level Examination (HKALE).

Until 2009, Hong Kong's education system was modelled on the British system. Secondary school education operated on a 5 + 2 model with the HKALE administered from 1979 to 2012 by the HKEAA at the end of Secondary 7 (Year 13), originally as The University of Hong Kong's entrance examination. In 2009, free compulsory education was extended to the end of Secondary 6 (Year 12), and from 2012, the HKALE was replaced by the HKDSE, which is now administered at the end of Year 12. Examination development and administration procedures for the HKDSE and HKALE are broadly similar. Development committees are run by the HKEAA for each subject, and comprise teachers and academics from the particular subject area.

Previous studies (eg, Coniam & Yeung, 2010; Coniam, 2011) have reported that, in the context of OSM in Hong Kong, markers had a positive perceived ease of use in the OSM environment and showed an increasing but slow acceptance of OSM. While marker reactions to OSM have been reported from a number of perspectives in previous studies, the potential difference in the attitude of markers in the context of specific subject areas is a macro-level picture for which, until now, there have not been sufficient data to make valid comparisons. The current study aims to fill this gap.

The first issue that might have an influence on marker reactions to OSM across different subject areas is the make-up of the different subject areas in terms of predominant (if any) question type. The questions on the different papers of the different subject areas in the HKDSE and HKALE include a mix of question types. Almost all subject areas involve extended response (ie, paragraph or essay questions), structured response (writing one or two lines) and limited response (single word, multiple choice) questions. The exceptions are mathematics, where the majority of the questions are limited response, and liberal studies, where the sole question type is the extended essay response. In the Yan and Coniam (2013) study, markers who marked limited response questions showed a higher level of perceived ease of use and acceptance of OSM than markers who marked extended essays—although significance only emerged regarding acceptance of OSM.

Another subject area-related issue involves the characteristics of the marker cohort for different subject areas and how these may contribute to differences among markers in terms of perceived ease of use and acceptance of OSM. It is not perhaps surprising to find different attitudes towards new technology, like OSM, among teachers of different subject areas given that they have different educational backgrounds and subject content areas. For example, Ball and Levy (2008) found that information systems instructors had a higher computer self-efficacy than instructors of other subject areas, and subsequently had a higher level of acceptance of emerging educational technologies than instructors of other subject areas. Given that Ball and Levy's (2008) study was based on a small sample ($n = 56$) of instructors from one university, their findings should therefore be interpreted with appropriate caution. Because OSM markers in Hong Kong are, in the

main, serving teachers, the aforementioned subject effect nevertheless warrants further exploration in the current study.

As practically all papers in all subject areas of both the HKDSE and HKALE public examinations were marked on screen, the current dataset comprises data from markers of subject areas from both examinations (see Appendix A). Similar subject areas in the two examinations are broadly comparable as are the question types (extended, structured or limited response questions, etc). Some markers (e.g., for English Language and Chinese Language) marked scripts from both HKDSE and HKALE examinations. Given that this scenario and that the majority of markers in many subject areas did not indicate whether they had been marking HKDSE or HKALE scripts, each subject area is henceforth treated as a single entity rather than differentiating between HKDSE and HKALE examinations. Data from both examinations have been included to present as full a picture as possible of markers for the 2012 examinations, the first time that all subjects had been marked in Hong Kong on screen.

Research questions

In the current study, the major focus involves comparing marker reactions from the wider perspective of the subject area. The research questions can be framed as:

1. What are the effects of subject area on markers' perceived ease of use in the OSM environment?
2. What are the effects of subject area on markers' perceived acceptance of OSM?

Methodology

There are three perspectives to the data in the current study. The first, and major, element of the data involves an analysis of the previously validated post-marking questionnaires, which markers responded to on a 6-point Likert scale. The second element of data is the open-ended sections of the questionnaire, where markers were asked to provide qualitative comments (positive or negative as they felt appropriate) about whatever aspects of the OSM system they wished to comment on. The third element—as triangulation against some of the results—involved post-hoc interviews with a key member of the staff from the HKEAA with a view to providing possible interpretations for some of the results that had emerged. These three perspectives will now be described in turn.

Responses to the Likert-type questionnaire

Sample

There is considerable variation in the size of the marking panels for different subject areas. Some subject areas—such as physical education and integrated science—have quite small candidates and hence only utilise a small number of markers. The number of papers in a given subject area is generally two. The exceptions are the compulsory English and Chinese language examinations, which have three or four different written papers and which consequently require a large number of markers.

Post-marking questionnaires were given to all 4551 markers who were involved in the OSM of the subject areas in 2012. A total of 1819 valid post-marking questionnaires were returned by markers for these 20 subject areas—a return rate of 40.0% (see Table 1).

Return rates varied considerably across different subject areas—possibly because of whether questionnaires were returned by post or in person. Some subject areas had a very high return rate: for mathematics, the return rate was 82.8%—a very high return rate for a marking panel with over 200 markers. Chinese language and culture (CLC) had the lowest return rate of 16.3%—possibly as a result of their having been canvassed through numerous questionnaires since the inception of OSM in 2007. The return rates figures can nonetheless be considered acceptable according to the 20–30% benchmark return rate reported for postal surveys without

Table 1: Marking panels and questionnaire return rates

<i>Subject area</i>	<i>Markers on panel</i>	<i>Questionnaires returned</i>	<i>Return rate (%)</i>
1. Liberal studies	1111	369	33.2%
2. English language	823	289	35.1%
3. Chinese language	663	224	33.8%
4. Mathematics (compulsory + extended)	232	192	82.8%
5. Chemistry	219	99	45.2%
6. Information and communication technology (ICT)	200	82	41.0%
7. Business, accounting and financial studies (BAFS)	154	75	48.7%
8. Geography	147	74	50.3%
9. Chinese history	127	74	58.3%
10. Biology	123	70	56.9%
11. Economics	120	69	57.5%
12. Physics	100	61	61.0%
13. Chinese language and culture (CLC)	251	41	16.3%
14. History	91	24	26.4%
<i>Sub-total</i>	<i>4361</i>	<i>1743</i>	<i>40.0%</i>
15. Chinese literature	100	20	20.0%
16. Computer applications and computer studies	26	19	73.1%
17. Physical education	15	14	93.3%
18. Ethics and religious studies	11	9	81.8%
19. Integrated science	16	8	50.0%
20. Health management and social care	22	6	27.3%
<i>Total</i>	<i>4551</i>	<i>1819</i>	<i>40.0%</i>

incentive (see eg, Blumberg, Fuller & Hare, 1974; Denscombe, 1998)—the situation in the current study.

Given that a sample size of 30 is generally accepted as the threshold for conducting statistical analysis (Ramsey, 1980), subjects in the current study with less than 30 respondents (physical education and integrated science, for instance) have been excluded from the analysis. History ($n = 24$), however, has been retained in the analysis as its sample size was only marginally below the target of 30 and it is considered a major subject. Table 1 presents details of marking panel sizes and questionnaire return rates for the 14 major subject areas used in the current analysis. The major subject areas—liberal studies, English language, Chinese language and mathematics—account, not surprisingly, for approximately 62% of respondents. The figure of 1743 markers, representing a return rate of 40.0%, across the 14 major subject areas represents a slight reduction on the optimum—that is, if all 20 subject areas were included in the analysis. The current dataset, however, comprises a sufficiently broad set of subject areas for a reliable and informed across-the-board picture to be presented.

Instrument

A recent study (Yan & Coniam, 2013) reported the validation of two scales, namely, *Ease of Use in the OSM Environment* and *Acceptance of OSM* using Rasch analysis (Rasch, 1960), with a view to laying the groundwork for the analysis of this final piece of the OSM jigsaw puzzle—a composite analysis of all subject areas. Two robustly calibrated scales consisting of a total of 13 items were determined by the Rasch analysis, with 7 items in the *Ease of Use in the OSM Environment* scale and 6 items in the *Acceptance of OSM* scale. The *Ease of Use in the OSM Environment* scale aimed at tapping markers' general computer proficiency, their competence in manipulating the mouse and enlarging and scrolling through the screen image, as well as ergonomic issues such as desktop height and screen resolution. The *Acceptance of OSM* scale was used to gauge issues such

as the accuracy of their onscreen/on-paper marking, how tired their eyes became through marking in the two modes and how often they needed to take a break while marking. It also enquired about their preference for marking mode, ie, OSM or paper-based marking. All items were posed on a 6-point Likert-type scale; response options ranged from “6,” a positive response or agreement, to “1,” a negative response or disagreement. Questionnaires were originally drafted (and piloted) in English as the majority of the markers would complete the questionnaires in English. Chinese versions of the questionnaires were, however, produced for subject areas whose candidates must answer in Chinese—Chinese language, CLC, Chinese history and Chinese literature. The translation–back-translation process was conducted by bilingual members of the research team to ensure linguistic and functional equivalence between English and Chinese versions of the questionnaire.

Open-ended comments

As mentioned, in addition to the Likert-scale questions, markers could also provide written comments to an open-ended question. Markers were asked to comment on any advantages and/or disadvantages that they perceived in marking on screen as opposed to on paper.

Around 65% of comments received were written in English and 35% in Chinese. All comments from markers for Chinese language, CLC and Chinese history were written in Chinese. As with the questionnaires, a bilingual member of the research team translated the comments into English, after which they were checked for veracity by a further bilingual researcher. Following this, comments were then categorised into main and sub-themes, with commonalities drawn out—principally with regard to perceived advantages and disadvantages of OSM—and frequencies tallied. In order to avoid overwhelming readers with data, frequency counts of markers’ comments are not provided for all subject areas. Only relevant aspects of the data are drawn upon—where it is felt these shed light on certain aspects of the data.

Interviews with key HKEAA personnel

With a view to providing insights into certain aspects of the data, a senior member of the staff in the Assessment Development Division of the HKEAA was interviewed, in order to triangulate aspects of the data for which the research team did not have adequate background information. This officer was a general manager who had been with the HKEAA over 20 years and who had a substantial role in the original planning and subsequent implementation of OSM. Prior to being interviewed, the HKEAA informant was sent the data and the analyses, along with a set of questions and queries that the research team wanted to probe. Related issues here included marker background, the composition and administration of OSM marking panels, and the time frame for the implementation of OSM in different subject areas. The ensuing semi-structured interview, which was conducted in English, lasted for about 90 minutes. Notes were taken during the interview, following which key points were highlighted, and referenced against relevant aspects of the data.

Results and discussion

This section presents results from two perspectives. First, the Rasch analysis results on the psychometric properties of the two scales, ie, *Ease of Use in the OSM Environment* and *Acceptance of OSM* scales, are briefly presented. The second perspective involves an examination of the impact of subject area on markers’ perceived ease of use and acceptance of OSM. While the major focus of analysis of this factor is predicated on the two scales from the post-marking questionnaire, the other two sources of data referred to above are used for triangulation. These are, one, markers’ responses to the open-ended sections of the questionnaire regarding markers’ perceived advantages and disadvantages of OSM; and two, the information and insights obtained from the interview with the HKEAA informant.

The psychometric properties of the two scales

Given that the major data resource is marker responses to the two scales, *Ease of Use in the OSM Environment* and *Acceptance of OSM*, the two scales need to function efficiently for valid comparisons to be made among subject areas. According to Messick's (1995) assertion that validity is an evolving property and validation should be a continuing process, the psychometric properties of the scales were examined based on the data in the current analysis from a Rasch measurement perspective. The criteria for judgement included category functioning, Rasch reliabilities, item fit statistics and variance explained by the measures. The results showed that a five-category rating scale for item 6 in *Acceptance of OSM* scale and a six-category rating scale for other items functioned well. Threshold calibrations advanced monotonically with category and no *threshold disordering* (Bond & Fox, 2007; Linacre, 2002) appeared. The Rasch person reliability was .85 for both scales, with item reliabilities of .99 and 1.00 respectively. Rasch measures were able to explain 50.4% and 63.5% of variance in the observed data for the two scales respectively, indicating that the Rasch model provides good predictions of both item difficulties and person measures. The mean square item fit statistics ranged from 0.83 to 1.39 for items in the *Ease of Use in the OSM Environment* scale and from 0.61 to 1.48 for items in the *Acceptance of OSM* scale, indicating an acceptable fit between the model and data. In summary, these two scales have good psychometric properties and are appropriate for use with the sample in the current study.

Cross-subject area comparison

The current section addresses the major focus of the study: viewing subject area as a single variable and comparing how markers as a group in each subject area differ on *Ease of Use in the OSM Environment* and *Acceptance of OSM*.

The data in this section are discussed from three perspectives. First, for each subject area, the picture of marker measures on the two scales is outlined. Second, an analysis of relevant written comments is presented. Third, discussions with the HKEAA informant are examined.

Marker measures on the two scales

With a view to moving the discussion towards the major issue of the current study, demographic-element fit statistics were checked by treating the subject area as a demographic facet and each subject area as an element within this facet in Rasch analysis. The results indicated that both scales showed good model fit. The mean square fit statistics for different elements (ie, subject areas) ranged from 0.83 to 1.21 for the *Ease of Use in the OSM Environment* scale and 0.77 to 1.21 for the *Acceptance of OSM* scale. Subsequently, the Rasch-calibrated marker measures were subjected to analysis of variance (ANOVA) with subject area as the explanatory variable. Table 2 presents a comparison of mean marker measures for the different subject areas. While the unit of the measurement scale is the logit, to aid interpretation of the results, "fair averages" (FAs) are also provided in parentheses, to the right of the logit measures. FAs (see Linacre, 1997, p 550 for details) are rating scale values converted from Rasch measures back to the original rating scale—the 6-point scale in this case. Such a conversion renders the output more easily interpretable by end-users, with the advantages of Rasch analysis maintained.

ANOVA results (see Table 2) indicated that the main effect of subject area for both scales was statistically significant ($p < .01$). To provide a more accessible comparative picture of the subject areas, Figures 2 and 3 provide graphical summaries of mean marker measures for different subject areas on the *Ease of Use in the OSM Environment* and *Acceptance of OSM* scales. In the figures, mean marker measures are represented by the horizontal lines, with vertical lines indicating plus or minus one standard error. The unit of the measurement scale is the logit, with FAs in parentheses.

Table 2: Mean marker measures for different subject areas

Subject areas	Number of markers	Ease of Use in the OSM Environment	Acceptance of OSM
Liberal studies	369	+1.70 (4.81)	-0.03 (3.60)
English language	289	+1.81 (4.87)	+0.06 (3.66)
Chinese language	224	+1.24 (4.56)	-0.04 (3.60)
Mathematics	192	+2.31 (5.09)	+1.09 (4.30)
Chemistry	99	+1.68 (4.80)	+0.35 (3.85)
Information and communication technology (ICT)	82	+2.80 (5.28)	+0.99 (4.24)
Business, accounting and financial studies (BAFS)	75	+1.56 (4.74)	-0.06 (3.58)
Geography	74	+1.13 (4.50)	-0.44 (3.34)
Chinese history	74	+1.42 (4.66)	+0.08 (3.67)
Biology	70	+1.19 (4.53)	-0.32 (3.42)
Economics	69	+1.50 (4.71)	-0.04 (3.60)
Physics	61	+1.85 (4.88)	+0.68 (4.05)
Chinese language and culture (CLC)	41	+1.55 (4.73)	+0.24 (3.78)
History	24	+1.10 (4.48)	-0.08 (3.57)
ANOVA results		$F(13, 1729) = 7.951,$ $p = .000$	$F(13, 1727) = 13.291,$ $p = .000$
Overall mean		+1.63 (4.76)	+0.18 (3.73)

ANOVA, analysis of variance; OSM, onscreen marking.

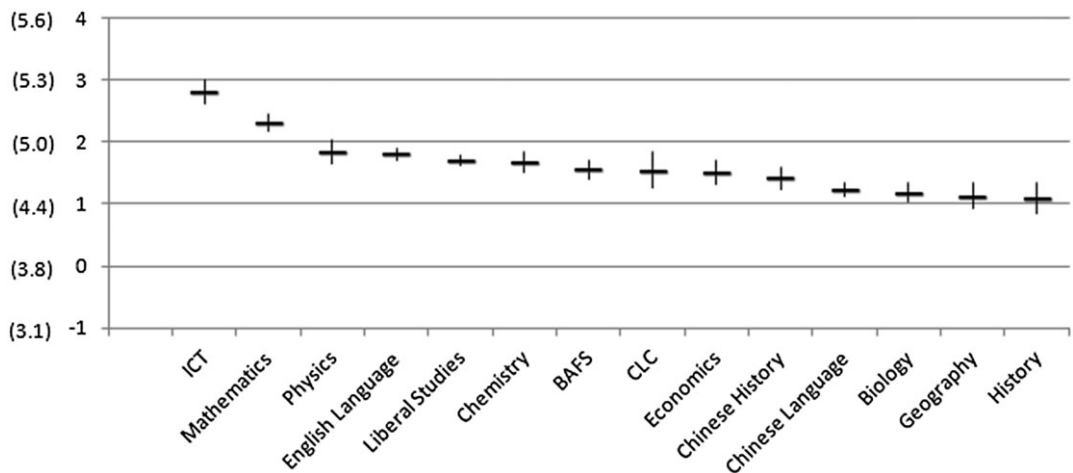


Figure 2: Mean marker measures for different subject areas on the Ease of Use in the OSM Environment scale. BAFS, business, accounting and financial studies; CLC, Chinese language and culture; ICT, information and communication technology; OSM, onscreen marking.

Whereas in Table 2 results have been presented in the format followed previously in this report of subject areas with the largest responses, or candidatures, being presented first, Figures 2 and 3 present the results from the perspective of most to least positive. This enables a clearer appreciation of comparisons across subject areas.

Figure 2 presents a comparison of mean marker measures for different subject areas on the Ease of Use in the OSM Environment scale. The adjusted FA value—corresponding to the zero logit—is 3.8, which is the psychological centre, or “consumer validity point” according to Coniam (2013).

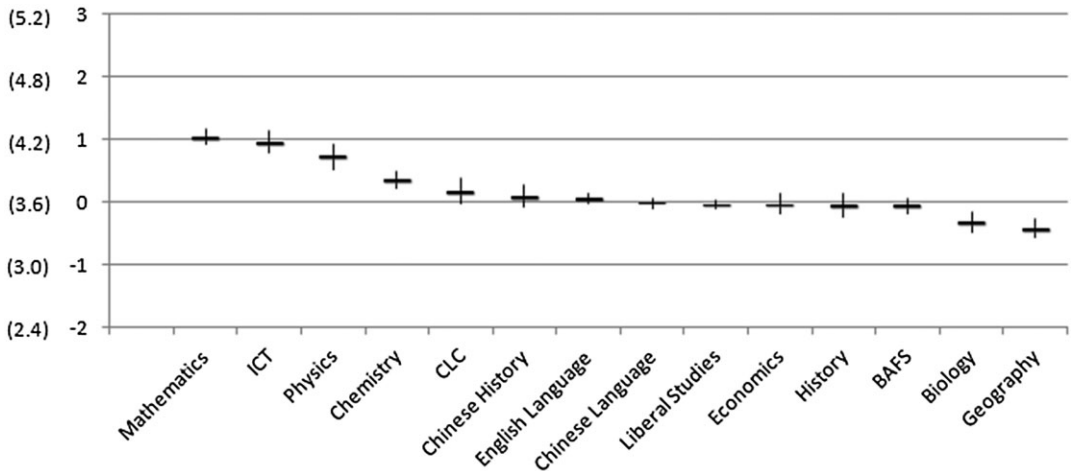


Figure 3: Mean marker measures of different subject areas on the Acceptance of OSM scale. BAFS, business, accounting and financial studies; CLC, Chinese language and culture; ICT, information and communication technology; OSM, onscreen marking.

As can be seen from Figure 2, markers for ICT (+2.80 logits, FA = 5.28), mathematics (+2.31 logits, FA = 5.09) and physics (+1.85 logits, FA = 4.88) exhibited the highest levels of perceived ease of use, with—possibly unsurprisingly—ICT recording the highest ease of use. While biology (+1.19 logits, FA = 4.53), geography (+1.13 logits, FA = 4.50) and history (+1.10 logits, FA = 4.48) showed comparatively less ease of use, all subject areas were nonetheless considerably above the zero point (a mid-point FA of 3.8) on the latent trait scale, indicating that markers for all subject areas were positive in terms of perceived ease of use.

Figure 3 now presents a comparison of mean marker measures for different subject areas on the *Acceptance of OSM* scale. On this scale, the mid-point FA value emerged at 3.6.

As Figure 3 illustrates, with a mid-point FA of 3.6, the majority of subject areas are at or above the “consumer validity” cutting point for acceptance of OSM. Markers for mathematics (+1.09 logits, FA = 4.30), ICT (+0.99 logits, FA = 4.24) and physics (+0.68 logits, FA = 4.05) also recorded the highest levels of acceptance of OSM (albeit with a slightly different order with mathematics just ahead of ICT). Biology (-0.32 logits, FA = 3.42) and geography (-0.44 logits, FA = 3.34) were in broadly similar positions on the chart to their locations on the previous ease of use scale, and with FA values below 3.6, demonstrating the lowest levels of acceptance of OSM. History (-0.08 logits, FA = 3.57) moved up the ladder slightly but was still towards the bottom of the acceptance scale. As can be seen from the two figures, the majority of subject areas appear in broadly similar positions on the *Ease of Use* scale (Figure 2) as they do on the *Acceptance* scale (Figure 3). The exception is business, accounting and financial studies (BAFS). This subject area appears in the middle of the *Ease of Use* scale (well above the mid-point of 3.8), yet towards the rear of the *Acceptance* scale, where it dips just below the 3.6 threshold.

In an analysis of marker severity in the Hong Kong-based *Language Proficiency of English Language Teachers* (LPATE) examination, Coniam (2008) suggests that half a band is a critical level of significance. Tests on the LPATE are scored on a five-point scale. Candidates must achieve a level ‘3’ on all scales in a given test component (one ‘2.5’ score is permitted). More than one 2.5 score or a level ‘2’ on any scale results in failure. In this context, half a band of difference can be seen as critical. While levels of acceptance of biology (-0.32 logits, FA = 3.42) and geography (-0.44 logits; FA = 3.34) clearly fall below the mid-point, they are not half a level below the threshold, indicating that the situation is not as serious as the impression in Figure 3 might suggest.

To conclude this key section regarding subject area, it is clear that in the *Ease of Use in the OSM Environment* scale, the positive perceptions by markers in all subject areas indicate clear confidence in being able to cope with the technical side of the demands of OSM. The mean marker measure for the *Ease of Use* scale was 1.63 logits. The FA mean of 4.7 was not only well above the mid-point of 3.8, but the scores for all subject areas were above this threshold. Regarding the *Acceptance of OSM* scale, the overall subject area mean was 0.18 (an FA mean of 3.73)—a positive indicator of acceptance. While a couple of subject areas were below the consumer validity mid-point, in the overall context of all subject areas, there would appear to be little cause for concern.

Given that there are differences between subject areas, the two following sections investigate issues in an attempt to account for such differences. The first reports markers' written comments; the second reports an analysis of the interviews with the HKEAA informant regarding factors related to different aspects of the make-up of different subject areas.

Analysis of written comments

As mentioned, markers were invited to write comments on any perceived advantages and disadvantages of OSM. These comments were consequently analysed with a view to investigating whether they provided any insight into the different reactions of markers of different subject areas. The comments were categorised into two major themes, ie, perceived advantages and disadvantages of OSM, and frequencies were tallied. Table 3 presents—generally for both the most and the least positive subject areas—the proportion of perceived disadvantages to perceived advantages by markers of ICT, mathematics and physics versus markers of history, geography and biology. Around 80% of comments from markers for these six subject areas were provided in English and 20% in Chinese.

As can be seen from Table 3, ICT, mathematics and physics markers provided considerably more advantage than disadvantage responses. Markers for history, geography and biology, in contrast, saw more disadvantages than advantages with OSM.

In order to examine whether the number of advantage/disadvantage entries provided by markers of different subject areas was significantly different, markers for ICT, mathematics and physics were combined into an "IMP" group (a *most positive* group) while markers for history, geography and biology were combined into an "HGB" group (a *least positive* group). Table 4 presents the total number of comments and the percentages of advantage comments for these two groups. The chi-square statistic revealed that the difference in the number of advantage/disadvantage entries between these two groups was statistically significant ($\chi^2 = 21.380, p < .01$).

As potential reasons for markers' generally positive attitudes towards OSM, samples of perceived advantages mentioned by the most positive group, ie, markers of ICT, mathematics and physics,

Table 3: Advantages perceived by markers for different subject areas

Subject area	Total comments	Perceived advantages (%)
ICT	107	62 (57.9%)
Mathematics	505	295 (58.4%)
Physics	188	100 (53.2%)
IMP group subtotal	800	457 (57.1%)
History	71	34 (47.9%)
Geography	220	99 (45.0%)
Biology	174	70 (40.2%)
HGB group subtotal	465	203 (43.7%)

ICT, information and communication technology.

Table 4: *Advantages perceived by most and least positive different subject areas*

<i>Subject area</i>	<i>Total comments</i>	<i>Perceived advantages (%)</i>
IMP/"most positive" group	800	457 (57.1%)
HGB/"least positive" group	465	203 (43.7%)

HGB, history, geography and biology; IMP, information and communication technology, mathematics and physics.

are provided. The most frequently mentioned advantages were related to the efficiency of OSM compared with paper-based marking. This point was acknowledged by 10 markers for ICT (12.2%), 40 markers for mathematics (20.8%) and 10 markers for physics (16.4%) by comments like:

- It's more efficient by using OSM (ICT).
- OSM reduces time to sort paper/flip paper (mathematics).
- I mark faster on screen than on scripts (physics).

The second most frequently mentioned advantage was the convenience caused by the freedom from having to handle the hard copies of scripts. Twelve markers for ICT (14.6%), 36 markers for mathematics (18.8%) and 7 markers for physics (11.5%) reported this point by stating "No need to carry bulky scripts," "Presenting scripts on screen is logistically convenient," etc.

To exemplify the views of the generally less positively oriented respondents, samples are provided below of disadvantages mentioned by markers of history, geography and biology. The most frequently mentioned issues related to the convenience of accessing the marking centre. Thirteen markers for history (54.2%), 44 markers for geography (59.5%) and 34 markers for biology (48.6%) reported this as a challenge with statements such as "too much travelling time," "need to go to the venue at a specific time," "too few centres," "difficult to book marking station," etc. Some typical comments were:

- Time consumed in traffic! (Biology)
- It is very inconvenient to travel to a special marking centre and do marking at designated time slots (geography).
- After a whole day's hard work at school, I have to spend almost one to two hours to travel to and from the marking centre (geography).
- Add a marking centre in Kowloon East (there are too few seats in the San Po Kong centre) (biology).
- Lengthen the opening hours of the marking centres. It's not easy to book work station (history).

In some ways, this finding is positive because it shows concerns about the logistics of the marking process rather than the marking itself. It is possible that some measures could be considered such as having more marking centres, as well as possibly running centres on a more flexible schedule so as to optimise centre usage. Another potential solution, suggested by some markers, involved an online marking system instead of physically fixed marking stations. Typical comments here were:

- Raise the level of security so as to allow markers to work online at home (mathematics).
- Online marking is recommended. No need to go to centre! (geography)
- Is it possible to set up a mini-marking station at our school? It would be a great relief if we would save the travelling time to and from the assessment centres (geography).

The second most frequently mentioned disadvantage was "eyes getting tired," which was mentioned by 11 markers for history (45.8%), 20 markers for geography (27.0%) and 17 markers for biology (24.3%). As one marker put it:

- Onscreen marking is not good to markers' health. After working on the screen for several hours, the eyes will be very tired and hurt (biology).

The findings regarding convenience of accessing the marking centre and potential eye strain appear to be important issues regarding future developments to the OSM system.

Interview with HKEAA personnel

As mentioned, a key HKEAA member of the staff was interviewed with a view to seeing how far her overarching expertise across different subject areas might help account for differences between subject areas. Three key points emerged from the interview.

The first area for consideration revolved around how different subject areas divide up their marking panels in terms of whether markers mark the entire paper (as was the practice in the past), whether they mark by section or by individual question. In this regard, subject areas differed in terms of how they assigned specific questions to their marking panels. Whether they assigned individual questions or sections to a "team" of markers tended to depend on the question types predominating in the examination paper. In general, subject areas involving extended response (ie, paragraph or essay questions) questions were marked by question, while subject areas involving structured response (writing one or two lines) and limited response (single word, multiple choice) questions were marked by section. Interestingly, this information regarding how marking panels were divided up reflected patterns of acceptance. For subject areas such as ICT, mathematics and physics that demonstrated the highest levels of perceived ease of use and acceptance of OSM, the examination papers were marked by section, while for subject areas exhibiting the lowest levels of perceived ease of use and acceptance, such as history, geography and biology, marking panels were based on individual questions. It may be the case that while marking by question is more efficient (see Coniam & Yeung, 2010), such a method of marking tends to be more monotonous than marking by section, where there is some degree of variety in the content—despite the fact that marking limited response questions is more mechanical. The marking of essay questions also means reading lengthy answers on screen, which may produce more "eye strain." Marking limited response questions also creates a better sense of progress as marks are input fairly frequently after each response is judged, whereas for longer essays, the marker needs to read a lot before inputting a mark, which may seem less "satisfying." Another possible contributing factor is related to annotations. Crisp and Johnson (2007) reported that annotations during marking not only relieve markers' administrative workload of marking, eg, keeping a running tally of marks, but also support higher order reading comprehension processes, especially for subjects involving longer answers. Although annotations are allowed in OSM systems, it is currently less easy for markers to annotate in an OSM environment than on paper scripts (Liu & Huang, 2008) and this may lead to less satisfaction with marking longer essays using OSM.

The second area discussed related to the differing lengths of time that subject areas have been implementing OSM. Certain subject areas, eg, English language and Chinese language, began implementing OSM early in 2007. Since then, OSM has been implemented progressively for various other subjects. OSM was implemented for ICT in 2008, for mathematics and physics in 2009 and for geography and biology slightly later—2010. For certain subject areas—economics, BFAS and chemistry, for example—2012 was their first time. The difference in length of OSM experience, however, demonstrated no direct impact on markers' perceived ease of use. Longer experience of OSM did not necessarily translate into higher levels of perceived ease of use/acceptance or vice versa: English language and Chinese language markers had been implementing OSM since 2007 yet were only mid-scale acceptors.

Another interesting observation by the HKEAA informant was that markers for certain subject areas—she specifically mentioned ICT and mathematics—appear to have been more adept at

mastering OSM and in adapting to the new marking environment. This group of subject teachers had sometimes been observed offering help to other markers in terms of adjusting computer settings and solving technical problems when marking centre technicians were not available.

Conclusion

As has been stated, the present study investigated the impact of subject area on markers' perceived ease of use in the OSM environment and acceptance of OSM from the comparative perspective of a large sample of markers—the entire cohort of markers for the 2012 HKDSE and HKALE examinations—marking all major subject areas.

The psychometric properties of the two scales, namely the *Ease of Use in the OSM Environment* and *Acceptance of OSM* scales, and the appropriateness of these scales for use with the sample in the current study were first examined from a Rasch measurement perspective. These two scales demonstrated good category functioning, acceptable Rasch person and item reliabilities, a high proportion of variance explained by measures and satisfactory item fit statistics. The robust psychometric properties of the two scales, therefore, provided a strong basis for valid comparisons to be made between subject areas.

Subject area was found to have a significant impact on perceived ease of use and acceptance of OSM. Markers for some subject areas, namely, ICT, mathematics and physics had the highest levels of perceived ease of use and acceptance of OSM, while markers for geography, biology and history demonstrated the lowest levels of performance on both scales. Bearing in mind the fact that markers are actually serving teachers for those subject areas, this result is in line with previous studies. As mentioned, Ball and Levy (2008) reported that information systems instructors have a higher computer self-efficacy and accept emerging educational technologies with greater ease than instructors of other subject areas. This is not a surprising finding given that teachers of technology-related subject areas generally have higher levels of exposure to new technology because of their educational backgrounds and the nature of their work than do teachers in other subject areas. The difference of efficacy and attitude regarding technology might therefore help account for the different levels of perceived ease of use and acceptance of OSM across subject areas. An observation by the HKEAA informant was that markers for some subject areas (ICT and mathematics) were more adept at mastering OSM and in adapting to the new environment.

Another possible factor that may account for why subject areas differ is the different approaches taken to how marking panels are divided up. The markers for subject areas that were marked by section, such as ICT, mathematics and physics, demonstrated higher levels of perceived ease of use and acceptance of OSM than the markers for subject areas that were marked by single question, such as history, geography and biology. The ways in which marker panels are formed are closely related to the question type predominating in the examination paper. Subject areas involving extended response questions were normally marked by question, while subject areas involving limited response questions were marked by section. In this sense, the finding is in line with Yan and Coniam's (2013) study where they found that markers marking limited response questions had higher levels of perceived ease of use and acceptance of OSM than those rating extended essays (although, possibly on account of small sample size, the difference regarding perceived ease of use did not reach statistical significance). A potential explanation is probably related to the reported disadvantage associated with reading on screen compared with reading on paper. Previous studies (eg, Tuncer, 2012; Tuncer & Bahadır, 2014) have revealed that reading on screen was less efficient than reading on paper. Furthermore, reading on screen was found to be a more cognitively demanding task than reading on paper (Wästlund, Reinikka, Norlander & Archer, 2005) and such increased cognitive workload might result in lower levels of comprehension of a text (Mayes, Sims & Koonce, 2001). This effect was most significant when readers read longer texts than when they read short texts (Johnson *et al.*, 2012).

Further, markers in subject areas that were on the positive end of the two scales confirmed their positive attitudes to OSM in the written comments they provided to the open-ended sections of the questionnaire. These markers (ie, of ICT, mathematics and physics) saw more advantages than disadvantages, while comments from markers of geography, biology and history reflected more disadvantages than advantages. So the fact that certain subject areas are more or less positively disposed to OSM is an issue that the HKEAA might need to consider with regard to how OSM is presented in the future to markers of different subject areas. Inconvenience in accessing the marking centre and eye tiredness appeared to be the most frequently mentioned disadvantages. It is presumed that the HKEAA will, at some point, consider measures to make marking centres more convenient for markers (eg, having more marking centres and/or running the centres on a more flexible schedule) or even developing an online marking system that can avoid the constraints of physically fixed marking centres. Such online marking systems are being experimented with in certain parts of the world (see, eg, Harrison, 2013, June) and will doubtless be on the HKEAA's future agenda. Potential eye strain is also an issue to be borne in mind regarding future developments and improvements to the OSM system. This is an echo of previous studies (eg, Tuncer & Bahadır, 2014) that labelled eye strain as one of the disadvantageous consequences of onscreen reading. Improvements in screen resolution (the HKEAA already uses 22-inch monitors) will, however, in time help to mitigate eye tiredness as will the quality of scripts scanned into the system with subsequent improved legibility.

To conclude, it is clear that the majority of markers across all subjects showed a high level of perceived ease of use. Markers of most subject areas also reported a generally positive acceptance of OSM—although they continue to report that their eyes get tired through marking on screen. These findings have important implications because, as stated above, since 2012, almost all HKDSE examinations have been marked on screen. The comprehensive picture of OSM presented in the current study nonetheless provides a positive signal for the shift from paper-based marking to OSM in the Hong Kong context as well as for any jurisdiction where the majority of examinations are being marked on screen.

Statements on open data, ethics and conflict of interest

Data access statement

We, the authors of this manuscript, will make the data used for this manuscript available for use as per request in the pursuit of scientific advance.

Ethical declaration

We, the authors of this manuscript, declare that this research study complied with all ethical guidelines stipulated by The Hong Kong Institution of Education and was approved by that institution.

Conflicts of interest statement

We, the authors of this manuscript, certify that there is no conflict of interest with any financial organisation regarding the material discussed in the manuscript.

Acknowledgments

We would like to thank the Hong Kong Examinations and Assessment Authority – and in particular Christina Lee, the General Manager for Assessment Development – for support on the project regarding access to markers and for data collection.

References

- Ball, D. M. & Levy, Y. (2008). Emerging educational technology: assessing the factors that influence instructors' acceptance in information systems and other classrooms. *Journal of Information Systems Education*, 19, 4, 431–444.

- Blumberg, H. H., Fuller, C. & Hare, A. P. (1974). Response rates in postal surveys. *The Public Opinion Quarterly*, 38, 1, 113–123.
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Mahwah, N.J.: Erlbaum.
- Campbell, A. (2005). Application of ICT and rubrics to the assessment process where professional judgment is involved: the features of an e-marking tool. *Assessment & Evaluation in Higher Education*, 30, 5, 529–537.
- Chang, S. S. (2000, June 13). *The ministry of education will promote on-screen marking in Gaokao*. People's Daily Online. Retrieved October 5, 2014, from <http://www.people.com.cn/GB/channel5/30/20000613/101191.html>.
- Coniam, D. (2008). An investigation into the effect of raw scores in determining grades in a public examination of writing. *Japan Association for Language Teaching Journal*, 30, 1, 69–84.
- 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. (2011). A qualitative examination of the attitudes of Liberal Studies markers towards onscreen marking. *British Journal of Educational Technology*, 42, 6, 1042–1054.
- Coniam, D. (2013). The increasing acceptance of onscreen marking—the 'tablet computer' effect. *Journal of Educational Technology & Society*, 16, 3, 119–129.
- 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.
- Crisp, V. & Johnson, M. (2007). The use of annotations in examination marking: opening a window into markers' minds. *British Educational Research Journal*, 33, 6, 943–961.
- Denscombe, M. (1998). *The good research guide for small-scale social research projects*. Buckingham, UK: Open University Press.
- DRS Group. (2014, July 4). *DRS forges a partnership in Southern Africa to increase adoption of electronic marking in exams*. Retrieved July 28, 2014, from <http://www.drs.co.uk/about-us/drs-news/2014/07/04/drs-forges-a-partnership-in-southern-africa-to-increase-adoption-of-electronic-marking-in-exams>.
- First use of onscreen marking for Zhongkao in Qingdao—results release on 4 July (2014, June 20). *Zhongkao Website*. Retrieved October 5, 2014, from <http://www.zhongkao.com/e/20140620/53a3a3200f119.shtml>.
- Geranpayeh, A. (2011). The impact of online marking on examiners' behaviour. *Research Notes*, 43, 15–21.
- Harrison, A. (2013, June). *Growth of online marking "could improve reliability."* BBC News. Retrieved October 10, 2013, from <http://www.bbc.co.uk/news/education-22798009>.
- 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.
- 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.
- Linacre, J. M. (1997). Communicating examinee measures as expected ratings. *Rasch Measurement Transactions*, 11, 1, 550–551.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 1, 85–106.
- Liu, Z. & Huang, X. (2008). Gender differences in the online reading environment. *Journal of Documentation*, 64, 4, 616–626.
- Lu, N. (2006, June 26). *More accuracy and fairness—comparing onscreen marking with paper marking*. Xinhua Net. Retrieved October 6, 2014, from http://news.xinhuanet.com/newscenter/2006-06/26/content_4752443.htm.
- Luo, Y. H. & Liu, T. M. (2009). Review of onscreen marking research. *China Examinations*, 11, 34–37.
- Ma, S. Y. (2004). Review and reflection of onscreen marking. *China Examinations*, 7, 24–26.
- Mayes, D. K., Sims, V. K. & Koonce, J. M. (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, 28, 367–378.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 9, 741–749.
- Raikes, N., Greatorex, 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/conf/proceedingsetc/IAEA2000NRJGSS>.
- Ramakrishna, A., Navya Sree, B., Sri Harish, P., 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.

- Ramsey, P. (1980). Exact type 1 error rates for robustness of student's t-test with unequal variances. *Journal of Educational Statistics*, 5, 4, 337–349.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright.
- RM Results. (2014). *Our story*. Retrieved October 8, 2014, from <http://www.rmresults.co.uk/content/success-stories>.
- Tuncer, M. (2012). The effect of reading from the printed material and the projection screen on remembrance comprehension and attention skills of the teacher candidates. *International Journal of Social Science*, 5, 7, 695–705.
- Tuncer, M. & Bahadır, F. (2014). Effects of screen reading and reading from printed out material on student success and permanency in introduction to computer lesson. *Turkish online Journal of Educational Technology*, 13, 3, 41–49.
- Wästlund, E., Reinikka, H., Norlander, T. & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: psychological and physiological factors. *Computers in Human Behavior*, 21, 377–394.
- Yan, Z. & Coniam, D. (2013). Assessing the ease of use in the environment and markers' acceptance of onscreen marking: a Rasch measurement perspective. *Educational Research and Evaluation*, 19, 5, 461–483.
- 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.
- Yang, M. F. (1999, August 15). *Internet, promoting the revolution of testing—using on-screen marking technology for Gaokao in the Guangxi province*. People's Daily Online. Retrieved October 6, 2014, from http://www.people.com.cn/rmrb/199908/15/newfiles/wzb_19990815001038_4.html.

Appendix A: HKDSE and HKALE subject areas marked on screen in 2012

HKDSE	HKALE
*Chinese language	*Chinese language and culture (CLC)
*English language	*Use of English language
*Mathematics (compulsory)	
*Liberal studies	Liberal studies
Chinese history	Chinese history
Chinese literature	Chinese literature
Geography	Geography
Mathematics (extended)	Mathematics
Physics	Physics
Biology	
Business, accounting and financial studies (BAFS)	
Chemistry	
Economics	
Ethics and religious studies	
Health management and social care	
History	
Information and communication technology (ICT)	
Integrated science	
Physical education	
	Computer applications and computer studies

Note. * indicates compulsory subject area.