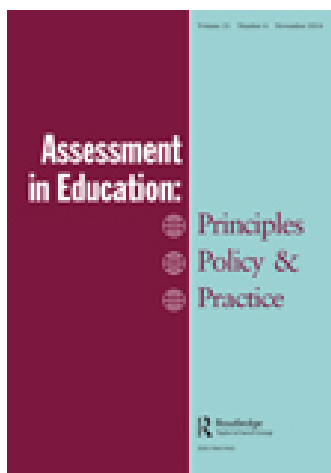


This article was downloaded by: [Hong Kong Institute of Education]

On: 06 October 2014, At: 19:38

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Assessment in Education: Principles, Policy & Practice

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caie20>

The effects of key demographic variables on markers' perceived ease of use and acceptance of onscreen marking

Zi Yan^a & David Coniam^a

^a Department of Curriculum and Instruction, The Hong Kong Institute of Education, Hong Kong

Published online: 02 Oct 2014.

To cite this article: Zi Yan & David Coniam (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](https://doi.org/10.1080/0969594X.2014.953910)

To link to this article: <http://dx.doi.org/10.1080/0969594X.2014.953910>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

The effects of key demographic variables on markers' perceived ease of use and acceptance of onscreen marking

Zi Yan* and David Coniam

Department of Curriculum and Instruction, The Hong Kong Institute of Education, Hong Kong

(Received 17 March 2014; accepted 6 August 2014)

The current study aims to investigate the effects of three key demographic factors – the language of marking, gender and age – on markers' reactions to onscreen marking (OSM). A total of 1743 markers completed a post-marking questionnaire consisting of two previously validated scales, i.e. *Ease of Use in the OSM Environment* and *Acceptance of OSM scales*. Rasch analysis results showed that the two scales had good psychometric properties. Markers generally reported finding the system easy to use and positive acceptance of OSM. Markers marking in both English and Chinese had higher perceived ease of use and acceptance than markers who marked only in English or in Chinese. Gender also had a significant impact on markers' responses to the two scales – favouring males. Age was not a significant factor influencing markers' perceived ease of use but older markers revealed a significantly higher level of acceptance than younger markers.

Keywords: onscreen marking; Rasch measurement; ease of use in the onscreen marking environment; acceptance of onscreen marking

Introduction

A new academic system was implemented in 2009 in Hong Kong. After completing three years of junior secondary education followed by three years of senior secondary education, all students take a single public examination (the Hong Kong Diploma in Secondary Education, or HKDSE) which aims to measure the attainment of students at the end of Year 12.¹ The HKDSE is a replacement for the previous Hong Kong Certificate of Education Examination (HKCEE) and the Hong Kong Advanced Level Examination (HKALE). The annual candidature for HKDSE in 2012 was 73,074. A major operational change accompanying the structural changes to the education and examination system is that onscreen marking (OSM) has been gradually replacing paper-based marking – to the extent that virtually all marking in the HKDSE in 2012 was conducted on-screen.² The current study furthers the research conducted into OSM in the Hong Kong context (see, e.g. Coniam, 2009a, 2009b, 2013; Coniam & Yeung, 2010; Yan & Coniam, 2013) in that the study's sample (as far as possible given constraints such as small sample size for several subject areas) includes all markers for all papers in all major subject areas marked on-screen in the 2012 public examinations.

*Corresponding author. Email: zyan@ied.edu.hk

As both qualitative and quantitative measures have illustrated (see Coniam, 2011; Coniam & Yeung, 2010), marker attitude has shown a slow but increasing readiness to accept OSM. While the current study encompasses the bigger picture in that it includes all markers, the micro-picture of certain key demographic factors, such as the language of marking, gender and age, which underpin marker reactions to OSM, has not been adequately explored in previous studies. These three key demographic variables are therefore investigated in the current study. To provide a background for the current study and to orient the reader, previous research regarding language issues in on-screen reading and marking, and the effects of gender and age on use of new technology will now be briefly outlined and the rationale for including these three variables in the current study justified.

Public examinations in Hong Kong can be answered in either English or Chinese – with the exception of English-related subject areas (English language, English literature) and Chinese-related subject areas (Chinese language, Chinese history, Chinese literature), for which answers must be in English or Chinese, respectively. Apart from English or Chinese, candidates are free to choose the language in which they wish to answer a subject area. Broadly, for science subject areas, approximately 50% or more of candidates opt to answer in English. For humanities subject areas, the figure is closer to 75% in Chinese. Regarding language of marking, markers also have the choice of the language they want to mark in. Most decide to mark solely in English or solely in Chinese, although some opt to mark in both languages.

A number of studies have reported salient differences between reading Chinese or English on paper in terms of speed and pattern (e.g., Rayner, Li, Williams, Cave, & Well, 2007; Sun, Morita, & Stark, 1985). With regard to these two languages in the context of reading on-screen, previous studies have indicated that the nature of the display appears to make a difference. For example, Chen and Chien (2005) investigated how the type of screen affected reading comprehension accuracy with smaller devices such as laptops and mobile phones. They found that reading accuracy can differ between English and Chinese depending on the display type. Tsai and McConkie (2003) pointed out that the writing system of English is an alphabetic system which uses a word as the perceptual unit, whereas Chinese uses a logographic writing system in which a character is used as the perceptual unit. Such differences lead to different characteristics of eye-movement between reading English and Chinese. Yen, Tsai, Chen, Lin, and Chen's study (2011) into the effects of typographic variables including character size, character spacing and font type on eye-movement measures in reading Chinese on-screen echoed Tsai and McConkie's (2003) argument concerning the apparently different characteristics involved in reading on-screen in Chinese or English. There are also differences in attitude towards on-screen reading emerging between English and Chinese readers. For instance, Liu (2005) reported a significant difference in preference for reading medium between American and Chinese readers. American readers showed a stronger reliance on paper as a reading medium than Chinese readers. Although cultural difference might be the major underlying reason for such differences, language could be another important contributor. The differences identified in previous studies in terms of reading Chinese or English on-screen justify the importance of researching the language issue in marking since markers have to read before marking.

The gender difference in terms of use of computers and new technology adaptation has long been a debatable topic, with previous studies generating often conflicting results. Zarrett and Malanchuk (2005) found that American adult males tended

to be more confident and interested in the use of computers and advanced technology skills than females. In a study on a sample of Indian college teachers, Suri (2003) reported that male teachers expressed more interest in computer technology and a higher level of confidence in the use of technology. Similarly, Belgian adult males were found to have more positive attitudes towards computers and the Internet, and less computer anxiety than females (Broos, 2005). A sample of male German college students (Sieverding & Koch, 2009) revealed higher computer self-efficacy for males than females. In an investigation into gender differences of Chinese undergraduates in the online reading environment, Liu and Huang (2008) found that male readers demonstrated a higher level of overall satisfaction with online reading than females, whereas female readers showed greater reliance on as well as stronger preferences for paper as a reading medium than male readers. A number of studies have nonetheless reported no significant gender differences in terms of interest, attitude and skills associated with the use of new technology. For example, Ahmad, Madarsha, Zainuddin, Ismail, and Nordin (2010) found that there were no significant gender differences regarding the actual use of technology for a sample of Malaysian college teachers. Sam, Othman, and Nordin (2005) also reported non-significant gender differences regarding interest in adopting technology and the skills demonstrated in technology use with a sample of undergraduates at a Malaysian university. Given that OSM is a new technology associated with computer use and there are inconclusive results regarding the effect of gender in new technology adaptation, it was felt necessary to include gender in the investigation on markers' attitudes towards OSM.

Age is another factor which has attracted substantial research interest, with studies focusing on attitudes towards the use of technology. Mixed results have been reported. For instance, while Teo (2008) found that age had no significant impact on attitudes towards computer use, Venkatesh, Morris, Davis, and Davis (2003) reported that it was more difficult for older users to accept new technology compared with younger users. Kennedy et al. (2008) also found that, compared with older participants, younger participants demonstrated more acceptance of technology by engaging more frequently in technology-based activities, such as the use of a phone as a personal organiser, retrieving information and downloading music from the Internet. The sample used in their study was, however, very large ($N > 2000$) – which could lead to very small differences being marked as significant. Consequently, since there was no substantial difference with practical implications between the mean scores of older and younger people on the use of these technologies, they concluded that the 'digital divide' between younger and older digital technology users was small. Similar to gender research, the mixed findings in terms of the role of age on the attitude towards and the use of technology indicate that further investigation is warranted.

A recent study (see Yan & Coniam, 2013) made an attempt to establish a robust scale for measuring markers' attitudes towards OSM in order to prepare the ground for the current extended study. Their study made use of three subject areas (English language, Chinese language and Liberal Studies) and successfully calibrated two scales, i.e. the *Ease of Use in the OSM Environment* scale (seven items) and the *Acceptance of OSM* scale (six items), from a Rasch measurement perspective (Rasch, 1960). Both scales showed satisfactory psychometric properties. The Rasch person reliabilities for the two scales are both .85. The Rasch person reliability is equivalent to traditional test reliability measures such as Cronbach's alpha. A value

of .85 indicates a satisfactory reliability of the scale. The *Ease of Use in the OSM Environment* scale tapped comfort issues in the OSM centre, markers' computer proficiency, their competency in manipulating the mouse, enlarging and scrolling the screen image, as well as ergonomic issues such as desktop height and screen resolution. These issues were identified as factors influencing markers' attitudes towards OSM. For example, in a study on online training, Elder, Barkhuizen, Knoch, and Von Randow (2007) found that technical features such as window space on the screen and ease of scrolling for information had an impact on markers' reactions to OSM. The *Acceptance of OSM* scale examined issues such as how accurate they felt their on-screen/on-paper marking was, 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 preferences for the mode of marking, i.e. OSM or paper-based marking. The scales are provided in Appendix 1.

The study

This section describes the data that make up the study, the research questions and the methodology used.

Data-set

While 2012 signalled the first marking of the HKDSE, it also marked the penultimate marking of the HKALE. The current data-set not only encompasses all HKDSE markers, it also includes HKALE markers of subjects that were marked on-screen. In relation to the objectives of the current study, whether marking was conducted for the HKDSE or the HKALE is not, however, an issue. Given that similar subject areas in the two examinations are broadly comparable, and that the majority of markers in many subject areas did not state whether they had been marking HKDSE or HKALE scripts, data and discussion about both examinations will be presented together without differentiating between the two examinations.

Research questions

The major research purpose of the current study involves investigating the extent to which – for the different languages used to mark, gender and age – markers' perceived ease of use in the OSM environment and acceptance of OSM differ. Specifically, the research questions are formulated as follows:

- (1) What are the effects of the language of marking (English, Chinese or both), gender and age on markers' perceived ease of use in the OSM environment?
and
- (2) What are the effects of the language of marking (English, Chinese or both), gender and age on markers' perceived acceptance of OSM?

Methodology

Instrument

As mentioned above, the questionnaire which markers completed after their marking comprised – in addition to their identifying marker number – two sections to gauge

(1) their perceived ease of use of OSM and (2) the level of their acceptance of OSM. All items were presented on a six-point Likert response scale, with (in accordance with Hong Kong Examinations and Assessment Authority [HKEAA] standard practice) '1' indicating a positive response or agreement, and '6' a negative response or disagreement. To aid interpretation, and to match Rasch measures (where positive values generally indicate 'more' of a trait and negative 'less'), the original codes were reversed so that '1' indicates a negative response or disagreement, and '6' a positive response or agreement.

Since the majority of the questionnaires would be completed in English, questionnaires were originally drafted (and piloted) in English. However, for subject areas which candidates must answer in Chinese – Chinese Language, Chinese Language & Culture, Chinese History and Chinese Literature – Chinese versions of the questionnaires were produced (after having been checked for consistency and veracity) by bilingual members of the research team.

Procedure

In the study, markers completed two short questionnaires. Prior to marking – at marker training sessions (March/April 2012) – all markers completed a pre-marking questionnaire. This questionnaire was in three sections and asked for demographic detail, assessment of markers' own technological proficiency, and their expectations towards the upcoming OSM exercise. Towards the end of the marking period (late May 2012), markers were provided with a post-marking questionnaire (the above-mentioned one), which they were asked to complete and return along with their marker report (a standard Hong Kong feature asking for their comments on the exam questions and candidate performance) after they had finished marking. 'Returning' may have involved physically dropping off the questionnaire at a marking centre, or posting it to the HKEAA using a prepaid envelope.

It should be noted that – in order to minimise inconvenience – markers were not asked to duplicate the demographic data that they had provided in the pre-marking questionnaire and the marker report. This would be retrieved from the system.

Sample

Post-marking questionnaires were given out to 4501 markers with 2496 markers returning a questionnaire. It was reported earlier that in order to minimise inconvenience, markers were not asked to duplicate the demographic data that they had provided in the pre-marking questionnaire and the marker report. In the results presented below, the data are compiled from markers who completed both a pre- and a post-marking questionnaire so as to identify the markers and to investigate the effects of demographic variables, which were provided in the pre-marking questionnaire, on markers' responses to the post-marking questionnaire. The restriction that the data on a given marker should contain both pre- and post-marking questionnaires reduces the size of the data-set – from 2496 to 1743, representing a response rate of 38.7%. The sample distributions for these three demographic factors – language of marking, gender and age – are presented in Table 1.

As can be seen from Table 1, the analysis of markers included in the current study indicates that 540 (31.0%) marked in English, 806 (46.2%) marked in Chinese while 235 (13.5%) marked in both languages. One hundred and sixty-two (9.3%) respondents did not provide any information in this regard. There were 978 (56.6%)

Table 1. Sample distributions for language of marking, gender and age.

Variable	Distributions	Number of markers
Language of marking	English	540 (31.0%)
	Chinese	806 (46.2%)
	English and Chinese	235 (13.5%)
	Missing	162 (9.3%)
Gender	Female	978 (56.6%)
	Male	751 (43.1%)
	Missing	14 (.8%)
Age	<26	16 (.9%)
	26–30	253 (14.5%)
	31–40	735 (42.2%)
	41–50	544 (31.2%)
	51–55	134 (7.7%)
	>55	43 (2.5%)
	Missing	18 (1.0%)

females, 751 (43.1%) males and 14 (.8%) who did not provide gender information. The majority of the markers fell into two age groups, with 735 (42.2%) in the 31–40 age group and 544 (31.2%) in the 41–50 age group.

Results and discussion

The Rasch analysis was undertaken using FACETS (Linacre, 2006b) with a grouped rating scale model by treating language of marking, gender and age as demographic facets (all elements being anchored at zero in the calibrations). This section presents results from three perspectives. First, the psychometric properties of the two scales, i.e. *Ease of Use in the OSM Environment* and *Acceptance of OSM*, are examined – from a Rasch measurement perspective – to ensure they are appropriate for use with the sample in the current study. Second, markers’ Rasch-calibrated measures on the two scales are obtained, after which the overall picture of markers’ perceived ease of use in the OSM environment and acceptance of OSM is presented. Third, the impact of the three ‘background’ demographic facets – language of marking, gender and age – on markers’ perceived ease of use and acceptance of OSM is investigated.

The psychometric properties of the two scales for use with the sample in the current study

The current study is predicated on the efficient functioning of the two scales: *Ease of Use in the OSM Environment* and *Acceptance of OSM*. Although these two scales have been validated in previous studies (Coniam & Yeung, 2010; Yan & Coniam, 2013), an investigation of the psychometric properties of the scales is necessary since validity is an evolving concept and validation is a continuing process (Messick, 1995). The items in these two scales need to function well for the sample used in the current study if valid investigations of the demographic variables are to be made. An analysis of item and scale quality, therefore, will first be presented.

The category functioning of the six-point Likert response scale was first checked. The results showed that there was no *threshold disordering* (Bond & Fox, 2007;

Linacre, 2002) for any items except for item A6, in the *Acceptance of OSM* scale, where threshold 3 (the intersection point between categories 3 and 4) had a lower calibration ($-.69$ logits) than that of threshold 2 (the intersection point between categories 2 and 3) ($-.57$ logits). This implied that category 3 was less likely to be observed and that this category could not consequently be considered a most probable response for any group of markers. Categories 2 and 3 were therefore combined. This resulted in a five-category rating scale for item A6. The resultant five-category rating scale for item A6 and the six-category rating scales for other items function well, with threshold calibrations advancing monotonically with category, indicating that higher response categories correspond to higher measures of the latent trait.

The two scales had acceptable Rasch person (both were $.85$) reliabilities for the sample used in the current study. Quite a high proportion of variance in the observed data (50.4% and 63.5% for these two scales, respectively) was explained by the Rasch measures – indicating that the Rasch model provides a good prediction of both item difficulties and person measures. Item fit statistics, including outfit and infit mean square (MNSQ), indicate the extent to which the data match the specifications of the Rasch model. The infit and outfit MNSQ of items ranged from $.83$ to 1.39 for the *Ease of Use in the OSM Environment* scale, and from $.61$ to 1.48 for the *Acceptance of OSM* scale. All item fit statistics fell into the range of $.5$ to 1.5 which has been suggested by researchers (e.g. Anshel, Weatherby, Kang, & Watson, 2009; Linacre, 2006a) as an indicator of acceptable model fit. For a more detailed picture of the results regarding the psychometric properties of these two scales, the reader is referred to Appendix 2. In summary, the results indicate that the two scales have good psychometric properties and are appropriate for use with the sample in the current study.

Markers' responses to the Ease of Use in the OSM Environment and Acceptance of OSM scales

The Rasch model calibrates person ('marker' in the current context) measures and item difficulty on the same latent trait scale. Appendix 3 presents the marker-item maps for the two scales. In the figures, the first column is the measurement scale shared by both markers and items. The scale origin – an arbitrary starting point on the Rasch scale like the end of a stick measuring length – is the mean of the item difficulties, i.e. zero logits. While the unit of the measurement scale is the logit, to aid interpretation of the results, 'Fair Averages' are also provided in parentheses, to the left of the logit measures. Fair Averages (see Linacre, 1997, p. 550, for more details) are rating scale values converted from Rasch measures back to the original rating scale – the six-point scale in the current study. Such a conversion renders the output more easily interpretable by end-users, with the advantages of Rasch analysis maintained. The second column presents markers' measures on each scale. Markers with higher levels of perceived ease of use/acceptance (higher scores on the scales) appear towards the top of the scale while those with lower levels of perceived ease of use/acceptance (lower scores on the scales) appear towards the bottom. The third column indicates item difficulty. Items with higher 'difficulty' levels (that is, on which markers tend to have low scores indicating negative responses) are placed towards the top, while those with lower difficulty levels (on which markers tend to have high scores indicating positive responses) are placed towards the bottom.

On the *Ease of Use in the OSM Environment* scale (Appendix 3, Figure 1(a)) the mean estimate of markers' perceived ease of use was +1.70 logits – considerably higher than the mean estimate of item difficulty (zero). The most difficult item to endorse was E5 (How comfortable were you reading off the screen?) and the easiest items were E2 (How would you rate your computer proficiency?) and E7 (How easily could you scroll the screen image?). The fact that the majority of markers had relatively high scores on all items – illustrated by their being located above the items (even the most difficult item E5) on the shared latent trait scale in the map – indicates that markers were generally satisfied with all aspects of the OSM environment, including comfort issues in the OSM centre (E1 and E5), markers' computer proficiency (E2), the ease of enlarging (E6) and scrolling (E7) the screen image, as well as ergonomic issues such as desktop height (E3) and screen resolution (E4).

On the *Acceptance of OSM* scale, the most difficult item to endorse was A1 (How tired did your eyes get marking on screen?), while the easiest item was A4 (Overall, how would you rate your onscreen marking experience?). Although the mean estimate of markers' acceptance of OSM was +.18 logits – close to the mean estimate of item difficulty (zero) – the majority of the markers were located above most of the items, except A1, on the latent trait scale (i.e. acceptance of OSM) (Appendix 3, Figure 1(b)). Markers generally viewed OSM in a positive light except for eye-strain.

Effects of language of marking, gender and age on markers' reactions to OSM

The major focus in the current study is the micro-picture of the effects of three demographic factors – the language of marking, gender and age which underpin marker response – on markers' perceived ease of use and acceptance of OSM.

Before conducting comparisons, facet fit statistics were examined to ensure that each facet fit the Rasch model. Language of marking, gender and age were treated as demographic facets with a view to obtaining demographic-element fit statistics. The results (see Table 2) showed that both scales had good model fit for the three

Table 2. Fit statistics for demographic facets (language of marking, gender and age).

	Ease of Use in the OSM Environment		Acceptance of OSM	
	Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ
<i>Language</i>				
English ($N=540$)	1.02	1.03	1.02	1.03
Chinese ($N=806$)	1.02	1.01	1.01	1.01
English and Chinese ($N=235$)	.93	.96	.96	.96
<i>Gender</i>				
Female ($N=978$)	.99	1.00	1.01	1.00
Male ($N=751$)	1.02	1.00	.99	1.00
<i>Age</i>				
<26 ($N=16$)	1.20	1.43	.91	.97
26–30 ($N=253$)	1.01	1.05	.85	.85
31–40 ($N=735$)	1.02	.99	1.03	1.03
41–50 ($N=544$)	1.00	1.00	1.02	1.01
51–55 ($N=134$)	.97	1.02	1.07	1.06
>55 ($N=43$)	.75	.76	1.01	1.01

demographic facets. The infit and outfit MNSQ of the elements in the three facets ranged from .75 to 1.43 for the *Ease of Use in the OSM Environment* scale and .85–1.07 for the *Acceptance of OSM* scale, indicating that all elements in the three facets fit the Rasch model for the two scales satisfactorily.

Since Rasch analysis transforms ordinal data into interval measures that are appropriate for traditional inferential statistics, the Rasch-calibrated marker measures on the two scales will be compared across different languages of marking (English and Chinese), male and female, and age groups. Tables 3–5 present comparisons of Rasch-calibrated mean marker measures for the three variables, commencing with Table 3 – the language of marking.

The results of ANOVA indicate that the main effect of language of marking for both scales was statistically significant ($p < .01$); *post hoc* comparisons indicated that the significant difference emerged with markers who marked in both English and Chinese. These markers had significantly higher ($p < .01$) scores (i.e. higher perceived ease of use and higher acceptance) than markers who only marked in one language – English or Chinese. The differences between markers who marked only in English or only in Chinese were, however, not significant.

It can be seen from Table 4 that gender appeared to be another important factor influencing marker responses to the two scales. T-test results indicated that males had a significantly higher ($p < .01$) level of perceived ease of use and acceptance of OSM than females. This confirms the findings reported in Coniam and Yeung (2010) for Liberal Studies where, on the issue of travelling to a marking centre (a necessity of OSM in Hong Kong), females were less accepting than males.

From Table 5, it can be seen that ANOVA for the two scales on the age variable showed that there were no significant differences among the different age groups in terms of perceived ease of use. However, in the acceptance of OSM, an interesting finding was that significant differences ($p < .05$) emerged favouring older markers (+.61 logits), with in general, the level of acceptance increasing with age.

Table 3. Comparison of mean marker measures across language of marking.

Language of marking	Ease of Use in the OSM Environment	Acceptance of OSM
English ($N = 540$)	+1.71 (4.85)	+09 (3.68)
Chinese ($N = 806$)	+1.61 (4.76)	+06 (3.66)
English and Chinese ($N = 235$)	+2.10 (5.00)	+85 (4.16)
ANOVA results	$F(2, 1578) = 7.534, p = .000$	$F(2, 1577) = 30.853, p = .000$

Note: Measures are in logits, with corresponding Fair Averages provided in parentheses to the right of the logit measures.

Table 4. Comparison of mean marker measures across gender.

Gender	Ease of Use in the OSM Environment	Acceptance of OSM
Female ($N = 962$)	+1.49 (4.70)	-.01 (3.62)
Male ($N = 743$)	+1.98 (4.95)	+44 (3.90)
t-test results	$t = -5.966, df = 1727, p = .000$	$t = -6.648, df = 1725, p = .000$

Note: Measures are in logits, with corresponding Fair Averages provided in parentheses to the right of the logit measures.

Table 5. Comparison of mean marker measures across age.

	Ease of Use in the OSM Environment	Acceptance of OSM
<26 ($N=16$)	+1.30 (4.60)	-.04 (3.60)
26–30 ($N=250$)	+1.68 (4.80)	+.01 (3.63)
31–40 ($N=724$)	+1.72 (4.82)	+.19 (3.75)
41–50 ($N=538$)	+1.66 (4.79)	+.17 (3.73)
51–55 ($N=130$)	+1.84 (4.88)	+.44 (3.90)
>55 ($N=43$)	+1.61 (4.76)	+.61 (4.01)
ANOVA results	$F(5, 1719) = .473, p = .797$	$F(5, 1717) = 2.500, p = .029$

Note: Measures are in logits, with corresponding Fair Averages provided in parentheses to the right of the logit measures.

Discussion and conclusion

The purpose of the present study has been to investigate the effects of the three key demographic factors (i.e. language of marking, gender and age) on markers' perceived ease of use in the OSM environment and acceptance of OSM in order to provide a detailed picture about markers' reactions to OSM.

The psychometric properties of two scales, namely the *Ease of Use in the OSM Environment* and *Acceptance of OSM* scales, were first checked via various statistics including: category functioning, Rasch person reliability, variance explained by measures and item fit statistics. The results showed that these two scales were psychometrically robust for use with the current sample and, therefore, gave credence to subsequent analysis.

On the whole, markers show a high level of perceived ease of use. In general, markers were satisfied with all aspects in the OSM environment. Markers reported a generally upbeat experience of OSM. Their overall acceptance of OSM was positive. These findings are in line with previous studies (e.g. Coniam & Yeung, 2010; Yan & Coniam, 2013).

Concerning the effect of the language of marking, results indicated that there was no significant difference between markers marking only in English and markers marking only in Chinese with respect to their perceived ease of use and acceptance of OSM (putting aside for the moment markers who marked in both languages). This finding is somewhat different from previous studies on a similar topic. Yan and Coniam (2013) found that markers marking in English had a significantly higher level of perceived ease of use and a higher level of acceptance of OSM than markers marking in Chinese. Possibly, the effect may be due to sample size: the number of teacher markers in the Yan and Coniam (2013) study was small with only very few teacher markers marking in English. The current sample is much larger, being more representative of the much broader gamut of subject areas.

As mentioned, markers who elected to mark in both English and Chinese (it will be recalled that markers indicated their preference for language of marking [English, Chinese, or both]) had significantly higher levels of perceived ease of use and higher acceptance than markers who only marked in English or only in Chinese. Given that there are different characteristics, such as eye-movement, involved in reading on-screen in Chinese or English (e.g. Tsai & McConkie, 2003; Yen et al., 2011), marking in both languages is apparently a more demanding task than marking in a single language. Since markers' perceptions of OSM task difficulty do not present a clear case as to why markers marking in both English and Chinese had higher levels

of perceived ease of use and higher acceptance, explanations need to be sought elsewhere – possibly in the personal characteristics of markers who chose to mark in both languages. One possible explanation, then, is that markers marking in both English and Chinese may be more flexible and willing to take on a challenge. Such greater flexibility may embrace acceptance of innovations such as OSM – echoed in a *post hoc* interview with a key HKEAA member of staff, who noted that markers who marked in both languages tended to be both very experienced teachers as well as very experienced and reliable markers. Consequently, she suggested that such markers were possibly more willing to rise to a challenge. Suto, Nádas, and Bell (2011) found that a marker's highest education is a significant predictor of marking quality. Whether there are meaningful correlations between markers' academic background, bilingual proficiency, marking quality and attitude towards OSM is, therefore, a worthwhile topic for future research. One possible implication – assuming that such markers can be identified – is that the HKEAA might wish to consider whether quasi-bilingual markers might be used wherever the opportunity arises.

Previous studies on the effect of gender on use of computers and new technology adaptation generated conflicting results. Some studies (e.g. Broos, 2005; Sieverding & Koch, 2009; Suri, 2003; Zarrett & Malanchuk, 2005) reported that males tend to develop more positive attitudes towards – and higher levels of confidence in – the use of new or advanced technology than females. In contrast, a number of studies (Ahmad et al., 2010; Sam et al., 2005; Teo, 2010) reported non-significant gender differences in terms of interest in adopting technology, attitude towards technology and the skills in the use of technology. Conceptualising OSM as an innovative, computer-based technology, the findings of the current study therefore align themselves with the former group of studies, with results showing that gender had a significant impact on markers' responses to these two scales. Males demonstrated a higher level of perceived ease of use and acceptance of OSM than did females. This finding is consistent with the results reported by Liu and Huang (2008) where male readers demonstrated a higher level of satisfaction with online reading than females. One possible explanation, as argued by Liu and Huang (2008), is that female readers tend to annotate more often when marking than male readers – obviously this preference cannot be accommodated in the current OSM environment. Another important contributor to this finding is probably the more positive attitude towards and higher level of confidence regarding the use of new technology by males, as reported by previous studies. Given that markers participating in the current study were serving teachers, the gender distribution across different subject areas might also contribute to the gender difference in terms of perceived ease of use and acceptance of OSM. In Hong Kong, the majority of teachers, and markers, in technology-related subjects (e.g. ICT, physics) are males who usually have higher levels of efficacy and more positive attitudes towards technology due to their greater exposure to new technology, educational background and the nature of their work.

Contrary to what has been reported in previous studies – that no significant difference exists among various age groups in terms of attitudes towards computer or technology adoption (e.g. Kennedy et al., 2008; Teo, 2008) or, if there were significant differences, younger groups were more positive in accepting new technology (e.g. Ahmad et al., 2010; Venkatesh et al., 2003) – the present study has revealed that, although age was not a significant factor influencing perceived ease of use in the OSM environment, it did impact upon markers' acceptance of OSM. Older markers tended to have higher levels of acceptance than younger markers. However,

a further examination of markers' responses to each item in the *Acceptance of OSM* scale revealed that the main effect of age was only significant ($p < .01$) for item A2 (How much pressure did you feel, knowing that your marking performance was being constantly monitored?). Older markers appeared to feel less pressure than their younger peers. In other words, the major source of difference in terms of acceptance of OSM came from markers' attitude towards being monitored during marking. It would appear that older markers, who will, in the main, be more experienced teachers and/or markers, have higher levels of tolerance for pressure during marking. The HKEAA might therefore profitably explore measures which may help lessen the pressure on (especially younger) markers, in particular with regard to monitoring techniques and feedback.

The current study has its limitations, however. The study primarily relied on quantitative survey data, except for the *post hoc* interview with a key HKEAA member of staff. This limited further investigation into the reasons underlying markers' responses. Without triangulating the in-depth qualitative data, a full interpretation and explanation of the survey results remains speculative. Future studies might seek evidence from other sources, such as markers' reflections, to re-examine or justify the findings of the study. Furthermore, the current study only focused on the effects of three demographic variables on markers' perceived ease of use and acceptance of OSM. Given that all papers in all major subject areas in public examinations have now been marked on-screen in Hong Kong from 2012 onwards, a potential area for future research could be the effect of different subject areas on markers' attitude towards OSM. Certain subject area-related issues – such as predominant question type (extended response questions or short answer questions) – as well as the interactions between the demographic variables and subject-related issues are worthwhile topics for investigation.

To conclude, the majority of markers clearly demonstrated quite a high level of perceived ease of use and were generally satisfied with all aspects of the OSM environment. Overall acceptance of OSM was therefore quite positive. Markers who chose to mark in both English and Chinese had significantly higher levels of perceived ease of use and higher acceptance than markers who only marked in English or only in Chinese. Male markers demonstrated a higher level of perceived ease of use and acceptance of OSM than did females. While age showed a significant impact upon markers' acceptance of OSM (favouring older markers), age was not a significant factor influencing perceived ease of use in the OSM environment. These findings have important implications for the improvement of the OSM system in Hong Kong by providing a comprehensive picture of marker reactions to OSM and the impact of demographic variables. As mentioned previously, the key demographic variables effectively underpin the bigger picture of subject areas as a whole. Research currently in train – to be subsequently reported on – involves examining the subject area as a variable, with a view to gauging how markers of different subjects perceive OSM issues.

Acknowledgements

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.

Notes

1. The HKDSE is taken by students at the end of six years of secondary education (Year 12); the first HKDSE was held in 2012. The HKALE was taken by students at the end of two-year sixth-form courses (Year 13); the last HKALE was held in 2013.
2. As from 2013, all HKDSE subjects are marked on-screen, with the exception of Literature in English Paper 2 where the marker has to make a marking decision after considering the coverage of the optional questions selected by the candidate.

Notes on contributors

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 Rasch measurement, scale development and large-scale assessment.

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.

References

- Ahmad, T. B. T., Madarsha, K. B., Zainuddin, A. M., Ismail, N. A. H., & Nordin, M. S. (2010). Faculty's acceptance of computer based technology: Cross-validation of an extended model. *Australasian Journal of Educational Technology*, 26, 268–279.
- Anshel, M. H., Weatherby, N. L., Kang, M., & Watson, T. (2009). Rasch calibration of a uni-dimensional perfectionism inventory for sport. *Psychology of Sport and Exercise*, 10, 210–216.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Broos, A. (2005). Gender and information and communication technologies (IT) anxiety: Male self assurance and female hesitation. *Cyber Psychology & Behaviour*, 8, 21–31.
- Chen, C.-H., & Chien, Y.-H. (2005). Effect of dynamic display and speed of display movement on reading Chinese text presented on a small screen. *Perceptual and Motor Skills*, 100, 865–873.
- Coniam, D. (2009a). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15, 243–263.
- Coniam, D. (2009b). Validating onscreen marking in Hong Kong. *Asia Pacific Education Review*, 11, 423–431.
- Coniam, D. (2011). A qualitative examination of the attitudes of Liberal Studies markers towards onscreen marking. *British Journal of Educational Technology*, 42, 1042–1054.
- Coniam, D. (2013). The increasing acceptance of onscreen marking – The 'tablet computer' effect. *Journal of Educational Technology & Society*, 16(3), 1–10.
- 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, 249–271.
- Elder, C., Barkhuizen, G., Knoch, U., & Von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing*, 24, 37–64.
- Kennedy, G., Dalgarno, B., Bennett, S., Judd, T., Gray, K., & Chang, R. (2008). Immigrants and natives: Investigating differences between staff and students' use of technology. In *Hello! Where are you in the landscape of educational technology? Proceedings of the Australasian Society for Computers in Learning in Tertiary Education, Melbourne, Australia* (pp. 484–492). Retrieved from <http://www.ascilite.org.au/conferences/melbourne08/procs/kennedy.pdf>
- Linacre, J. M. (1997). Communicating examinee measures as expected ratings. *Rasch Measurement Transactions*, 11, 550–551.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85–106.

- Linacre, J. M. (2006a). *A user's guide to WINSTEPS/MINISTEP: Rasch-model computer programs*. Chicago, IL: Winsteps.com.
- Linacre, J. M. (2006b). FACETS (Version 3.6) [Computer Software and manual]. Chicago, IL: Winsteps.com.
- Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past 10 years. *Journal of Documentation*, 61, 700–712.
- Liu, Z., & Huang, X. (2008). Gender differences in the online reading environment. *Journal of Documentation*, 64, 616–626.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Copenhagen: Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago, IL: University of Chicago Press.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47, 2714–2726.
- Sam, H. K., Othman, A. E. A., & Nordin, Z. S. (2005). Computer self-efficacy, computer anxiety, and attitudes toward the Internet: A study among undergraduates in Unimas. *Educational Technology & Society*, 8, 205–219.
- Sieverding, M., & Koch, S. C. (2009). (Self-)evaluation of computer competence: How gender matters. *Computers & Education*, 52, 696–701.
- Sun, E., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics*, 37, 502–506.
- Suri, S. (2003, August 20–23). *Attitudinal barriers amongst teachers in distance teaching institutions of India*. Paper presented at the Forum on ICTs & Gender: Optimizing Opportunities, Kuala Lumpur, Malaysia.
- Suto, I., Nádas, R., & Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education*, 26, 21–51.
- Teo, T. (2008). Pre-service teachers' attitudes towards computer use: A Singapore survey. *Australasian Journal of Educational Technology*, 24, 413–424.
- Teo, T. (2010). Establishing gender structural invariance of Technology Acceptance Model (TAM). *The Asia-Pacific Education Researcher*, 19, 311–320.
- Tsai, J. L., & McConkie, G. W. (2003). Where do Chinese readers send their eyes? In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 159–176). Amsterdam: Elsevier.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Towards a unified view. *MIS Quarterly*, 37, 425–478.
- 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, 461–483.
- Yen, N.-S., Tsai, J.-L., Chen, P.-L., Lin, H.-Y., & Chen, A. L. P. (2011). Effects of typographic variables on eye-movement measures in reading Chinese from a screen. *Behaviour & Information Technology*, 30, 797–808.
- Zarrett, N. R., & Malanchuk, O. (2005). Who's computing? Gender and race differences in young adults' decisions to pursue an information technology career. *New Directions for Child and Adolescent Development*, 110, 65–84.

Appendix 1. Onscreen marking questionnaire

Please enter your marker number and responses on the attached computer sheet. Fill in the ovals, or enter detail as appropriate.

Marker number:

Scale One: Ease of Use in the OSM Environment scale

E1. How comfortable was the marking area in the assessment centre (general ambience, space, lighting, air-con, etc.)?	very comfortable 1 2 3 4 5 6 very uncomfortable
E2. How would you rate your computer proficiency?	very proficient 1 2 3 4 5 6 not proficient at all
E3. Was the desktop at the right height for you?	exactly right 1 2 3 4 5 6 very bad indeed
E4. How was screen resolution?	very good 1 2 3 4 5 6 very poor
E5. How comfortable were you reading off the screen?	very comfortable 1 2 3 4 5 6 not comfortable at all
E6. How easily could you enlarge the screen image?	very easily 1 2 3 4 5 6 with much difficulty
E7. How easily could you scroll the screen image?	very easily 1 2 3 4 5 6 with much difficulty

Scale Two: Acceptance of OSM scale

A1. How tired did your eyes get marking on screen?	not tired at all 1 2 3 4 5 6 very tired
A2. How much pressure did you feel, knowing that your marking performance was being constantly monitored?	no pressure at all 1 2 3 4 5 6 a lot of pressure
A3. How helpful did you find the support and feedback from the OSM system?	very helpful 1 2 3 4 5 6 not helpful at all
A4. Overall, how would you rate your onscreen marking experience?	very good 1 2 3 4 5 6 very bad
A5. How do you now feel about the move from paper-based to onscreen marking?	a good move 1 2 3 4 5 6 a bad move
A6. Would you prefer to mark on screen or on paper?	on screen 1 2 3 4 5 6 on paper

Appendix 2. Item fit statistics, item measures with standard errors (in parentheses), and category thresholds

Scale/item	Infit MNSQ	Outfit MNSQ	Difficulty	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5
<i>Ease of use in the OSM Environment scale</i>								
E1	1.05	1.07	.06 (.04)	-3.26 (.42)	-1.00 (.14)	-.75 (.09)	.78 (.06)	4.22 (.09)
E2	1.34	1.39	-.40 (.04)	-3.52 (.72)	-1.06 (.17)	-.52 (.10)	.93 (.06)	4.17 (.08)
E3	.91	.88	-.13 (.04)	-2.98 (.52)	-1.59 (.17)	-.51 (.09)	.68 (.06)	4.40 (.09)
E4	.91	.89	-.28 (.04)	-3.13 (.59)	-1.26 (.18)	-.71 (.10)	.88 (.06)	4.22 (.09)
E5	.97	.99	1.04 (.03)	-2.66 (.18)	-1.48 (.10)	-.99 (.07)	.65 (.06)	4.48 (.14)
E6	.96	.96	.14 (.03)	-2.45 (.27)	-1.47 (.12)	-.67 (.07)	.80 (.04)	3.79 (.06)
E7	.86	.83	-.44 (.04)					
Rasch person reliability: .85								
Variance explained by measures: 50.4%								
<i>Acceptance of OSM scale</i>								
A1	1.05	1.04	1.56 (.03)	-2.78 (.07)	-.98 (.06)	-.34 (.07)	.65 (.11)	3.45 (.29)
A2	1.48	1.48	.05 (.03)	-2.71 (.13)	-1.30 (.07)	.21 (.06)	.85 (.07)	2.95 (.12)
A3	1.43	1.43	-.42 (.03)	-2.85 (.20)	-1.43 (.10)	-.74 (.07)	1.33 (.06)	3.68 (.13)
A4	.70	.68	-.84 (.03)	-2.62 (.23)	-1.09 (.11)	-.69 (.08)	.84 (.06)	3.56 (.10)
A5	.61	.61	-.52 (.03)	-2.17 (.15)	-.89 (.09)	-.72 (.07)	.81 (.06)	2.97 (.09)
A6	.74	.74	.16 (.03)	-1.75 (.08)	-.30 (.06)	.18 (.06)	1.87 (.09)	N/A
Rasch person reliability: .85								
Variance explained by measures: 63.5%								

Appendix 3. Marker-item maps

1(a) Ease of Use in the OSM Environment scale

Measr	+Marker	-Item
(5.9) 7	+	*
(5.8) 6	+	*
(5.7) 5	+	.
(5.4) 4	+	**
(5.1) 3	+	***
(4.8) 2	+	****
(4.4) 1	+	*****
(3.9) 0	*	*****
(3.0) -1	+	.
(2.4) -2	+	.
(1.9) -3	+	.
Measr	* = 21	-Item

1(b) Acceptance of OSM scale

Measr	+Marker	-Item
(5.8) 6	+	.
(5.7) 5	+	.
(5.5) 4	+	.
(5.3) 3	+	*
(4.9) 2	+	****
(4.3) 1	+	*****
(3.6) 0	*	*****
(2.8) -1	+	*****
(1.9) -2	+	*
(1.8) -3	+	.
(1.7) -4	+	.
(1.4) -5	+	.
Measr	* = 14	-Item

Downloaded by [Hong Kong Institute of Education] at 19:38 06 October 2014