

Online Proctoring of High-Stakes English Language Examinations: A Survey of Past Candidates' Attitudes and Perceptions

David Coniam¹, Leda Lampropoulou² & Angeliki Cheilari²

¹ Head of Research, LanguageCert, UK

² Head of Assessment, LanguageCert, UK

Correspondence: Leda Lampropoulou, Head of Assessment, LanguageCert, 121 Sloane Street, London, SW1 X9BW, UK.

Received: June 25, 2021

Accepted: July 19, 2021

Online Published: July 27, 2021

doi: 10.5539/elt.v14n8p58

URL: <https://doi.org/10.5539/elt.v14n8p58>

Abstract

This paper reports reactions by candidates to the use of online proctoring (OLP), 'invigilation', in the delivery of high-stakes English language examinations.

The paper first sets the scene in terms of the move from face-to-face to online modes of delivery. It explores the challenges and benefits that both modes offer, in terms of accessibility, fairness, security and cheating.

Evidence is then presented from a survey exploring the reactions to and perceptions of OLP by candidates who had taken an English language examination via OLP. A strong endorsement of OLP was generally recorded. Feedback revealed that respondents perceived OLP to be a more personal as well as a more efficient way of taking a test. Some pertinent negative comments from a smaller number of respondents could be construed as constructive and are also discussed. The results are indicative of a broad acceptance of OLP, pointing to strong future uptake of the OLP mode of test delivery.

Keywords: online invigilation, survey, English language assessment

1. Introduction

1.1 *Online Delivery of Learning and Teaching*

The mode for 'delivery' of both teaching and assessment has long been accepted as the teacher standing at the front of a class of students, to whom she provides input or whom she facilitates (Wiesenberg & Stacey, 2008). In line with developments in technology and its uptake and acceptance across all facets of society (Lim & Wang, 2016), views of how education may be delivered are changing, spurred on considerably by the 2020 Covid-19 pandemic (TPD@Scale for the Global South, 2020). The traditional mode of delivery is consequently diminishing as the use of more innovative and interactional methods grows – a clear example of which being the development of Massive Open Online Courses (MOOCs) over the past decade; see e.g., Bonk et al. (2015).

When it comes to the impact of Covid-19, Todd (2020) presents a cogent discussion of how it has forced teachers to consider – and to immediately handle – online teaching.

Over the past decade, developments in technology have fuelled a shift to a more 'blended' form of learning and teaching (see Lim & Wang, 2016) and as a result, many countries have moved to embrace, or are at least experiment with, different forms of synchronous and asynchronous modes of teaching (Lim & Graham, 2021).

While the world is perhaps undergoing a degree of what Williamson (2015) calls 'datafication', that is "the objective quantification of all kinds of human behaviour and sociality" (p. 124), this new paradigm in education is permitting greater access to knowledge and education by previously more marginalised sectors of society, as the work of Lim & Graham (2021) illustrates.

Nonetheless, while greater use of technology, especially in blended learning, has brought about a change in mindset in terms of the delivery of teaching content using online instruction and facilitation, the delivery of examinations is still largely expected to occur in a face-to-face situation (Ahlawat et al., 2014).

Assessment – high-stakes assessment in particular – has long been conducted in front of an examiner in a centre such as a school/university hall or in a computer laboratory under examination conditions. Following completion

of the test, the test paper is submitted to marker(s), and the result then emerges often after a considerable period of time. There has been some take-up of technology in the area of assessment, although less than with teaching. A brief examination of test administration permutations afforded by technology is presented in Table 1 below:

Table 1. Test administration permutations

<i>Test taker location</i>	<i>Invigilator location</i>	<i>Invigilation conducted</i>
at a centre	in person at a centre	by examiner
at home	remotely (via video link)	by invigilator
at home	–	by self

Alternative modes of test administration – in addition to the traditional – involve a test taker taking a pen-and-paper (or a computer-based) test in their own home, invigilated by an examiner (proctor) in another location (via video link) or a speaking test administered remotely by an examiner. Taking an exam unsupervised, as is the case with take-home exams (Bengtsson, 2019), is also sometimes the case.

Hussein et al. (2020) comment that, while typical learning and teaching may be conducted quite competently by current online learning technologies, the conducting of assessment is fraught with challenges and problems. A major issue centres around the ability to ensure academic integrity when examinations are taken remotely, and from a private location such as the test taker's home.

1.2 Online Delivery of Assessment

There are a number of issues surrounding the online delivery of assessment; some of which may be seen as positive and others negative.

The Covid-19 pandemic has forced a rapid evolution of how education is delivered, with many educational institutions moving extremely rapidly to partial if not total online delivery of classes – see Gardner (2020). Khan and Jawaid (2020) discuss the issue of technology enhanced assessment (TEA) during the Covid-19 pandemic in Pakistan. They conclude with the observation that it is not now possible to “shy away from online teaching, learning and assessment” (p. 3) – the key word here is *assessment*: Assessment needs a deep rethink, but it has faced much less of a sea change than learning and teaching. However, a major issue is the strong concern with security.

Clark et al. (2020) comment on the ‘fit’ of instructional practices within the course in terms of online versus face-to-face teaching and assessment. They comment on the issue of ‘continuity’ stating that where a course is intended at the outset to be a distance learning one, then online assessment should naturally fall into place. During Covid-19, while teachers managed to adapt to an extent to online teaching, assessment was still a bolt-on. The mismatch between intended course outcomes and assessment conducted online tended to be greater than with paper-based assessment which appears more aligned with intended course outcomes. Clark et al. (2020) suggest that if the gap between intended course outcome and assessment is to be narrowed, classes need to begin with a distance learning format. In this way, students will become acclimatised to such an environment and be prepared for taking online exams and will more clearly see the fit between online assessment and course content.

In Ardid et al.'s (2014) study of groups of students given online non-proctored exams, the participants scored higher than those who were given online proctored exams. They thus raise two key concerns in discussing how non-proctored assessments can be conducted satisfactorily: concerns about security and honesty.

Differences between proctored exams and proctored paper-and-pencil exams have been investigated by, for example, Alexander, et al. (2001). In the context of a computer technology course, the latter found no significant difference in student performance on proctored online exams and proctored paper and pencil exams.

In a study of 1,455 test takers of the Canadian Academic English Language (CAEL) test, Zumbo (2021) reported no significant differences for tests administered at a test centre or for tests administered online.

1.2.1 Advantages of Online Proctored Exams

On the positive side, test takers may take a test in the comfort (and safety) of their own home – an important factor in times of a pandemic where movements are restricted or, particularly, for a person who is disabled in some manner. Convenience and speed are another factor to be considered; a test may be delivered via computer, and results may therefore be obtained more quickly.

As stated earlier, a majority of assessments – high-stakes school and university tests for example – involve test takers sitting in a hall and writing by hand for two hours. Given that the majority of assignments which test

takers will have written over the course of an academic year will have involved multiple drafts via a computer word processor, it well may be argued that the traditional mode of administering exams actually compromises validity because the realities of traditional examination conditions do not reflect real life and are thus, in a sense, invalid (Mogey et al., 2012). The possibility of completing a test via the word processor on a locked-down computer actually offers a test taker a more valid mode in which to complete an examination.

1.2.2 Potential Drawbacks

One stumbling block concerns expectations of teaching outcomes vs. expectations of assessment outcomes. Online teaching strongly stresses collaborative principles, such as discussion, peer support, learning tailored to individuals, self-regulated learning, getting students to set their own goals, plan, monitor and control their cognition (Boekaerts & Corno, 2005). In contrast, expectations of online assessment (and in particular high-stakes assessment) are that the result will be generated by one test taker, working on their own, with no recourse to any form of external support. For comparability's sake, this generally means that the same test needs to be delivered to a given group of individuals at the same time. This requirement therefore impacts on security, honesty, fairness and reliability, which leads to considerations of test takers gaining an 'advantage', i.e., cheating.

1.3 Security

There has been considerable discussion in the literature about levels of security for different types of online examinations. Foster (2013) presents a cogent overview of security in online proctoring, which is useful as a lens through which high-stakes assessment can be viewed. Foster (ibid) defines online proctoring by putting emphasis on "the critical use of the Internet and automated processes to produce a secure solution in monitoring test takers" (p. 2), whereas he generally defines remote proctoring as the human-invigilation of examinations, often in lower-stakes situations.

Foster presents a comprehensive list of key security features, ranging from the management and training of the proctor to interaction with the test taker, to the stability of the internet, to data transfer encryption. The list is laid out in Table 2.

Table 2. Key Security Features in OLP examinations (after Foster, 2013)

<i>Key Feature</i>	<i>Subsumed under each feature</i>
1. Online proctor during exam	
2. Continuous internet	
3. Encryption for data transfer	
4. Schedule availability	
5. Proctor management	Supervised / training / career path /certification
6. Interaction with test taker	Live chat / canned messages / live instruction to examinee / proctor views examinee screen / proctor as collusion threat
7. Prevent proctor view of screen	
8. Later video review proctoring	
9. Later video review capable	
10. Control during test session	Test launch / pause test / suspend test / cancel test
11. Automated proctoring	Inappropriate keystroke / audio levels / real-time data forensics
12. Lockdown	
13. Authentication	
14. Webcam	
15. Logs/records	Video storage / session review / time-stamped incident / incident logs
16. Program customization	Levels of security decisions / allowed/specified aids
17. Effectiveness research	

Foster describes how systems can provide levels of security which make online proctoring of examinations viable. He comments on the disadvantages that may be associated with traditional proctoring, where the proctors

may be corrupt or may want to influence candidates scores in some way. Indeed, a number of studies report how exam security may be stronger as a result of the technologies associated with monitoring of online examination than in traditional face-to-face settings (Rose, 2009; Watson & Sottile, 2010).

LanguageCert has a rigorous set of regulatory principles for online delivered and online proctored examinations and assessment relating to test security, format and the personnel involved. These adhere closely to the guidelines and recommendation laid down by the UK's Qualifications and Curriculum Authority (2007) and predate Foster's (2013) set of security features provided in Table 2 above. To exemplify, upon first log-on, candidates need to follow a thorough 'onboarding' process; this includes an ID check, locking down their computer, checking there are no second monitors, and a room check through their webcam to show that the room is secure and that no other person or aids are present. The behaviour of candidates during the examination is then monitored in a number of ways: via qualified, specially trained proctors, and the use of video and audio recording, as well as advanced video and audio analytics and surveillance software.

1.4 Cheating and Academic Dishonesty

Cheating in exams is hardly a new phenomenon. Before the advent of the digital age and much easier access to the internet and plagiarism, comments about candidates cheating in examinations abound – see e.g., Wright & Kelly, 1974; Bushway & Nash, 1977; Sierles et al., 1980. The Carnegie Council Report (1979) some forty years ago made reference to a growing "ethical deterioration" in academic life in terms of the number of college students cheating to get their desired grades.

The internet has certainly brought issues of cheating more to the fore over the past decade, since it offers access to digital documents and networks of people willing to facilitate paid cheating (Harper et al., 2020). Cheating in online examinations is becoming more widespread, and has been explored in numerous studies (Harmon & Lambrinos, 2008; Grijalva et al., 2006; Watson & Sottile, 2010).

Corrigan-Gibbs et al. (2015) provide an extensive debate on the extent of cheating and academic dishonesty, emphasising on the vulnerability of online tests as a salient concern. Much of the research in this area focuses on ways of making online exams secure and discouraging or preventing cheating, and how some test takers attempt to undermine these efforts. However, as mentioned above, some researchers (Rose, 2009; Watson & Sottile, 2010) suggest that, compared with traditional face-to-face settings, online tests may be as, if not more, secure than face-to-face tests, provided that adequate protocols are in place.

1.5 Learners' Attitudes and Perceptions towards Online Learning and Assessment

With the greater uptake of online learning over the past two decades, there has been a considerable number of studies which have explored learners' attitudes towards the method and the medium – see e.g., Hos et al., 2016; Rahmawati, 2016; Cakrawati, 2017; Erarslan and Topkaya, 2017. In general, perceptions have been positive – as the studies above all report.

While the greater acceptance of online learning has understandably seen considerable exploratory research, this has not been the case for online assessment, which initially saw comparatively less uptake than did online learning. Although fewer studies have investigated online assessment, some key studies are nonetheless described below.

Ozden (2004) explored computer-assisted assessment, with participants reporting favourably on the use of ICT (information and computer technology) in assessment. In Dermo's (2009) survey of emotions experienced by learners during online assessments, results indicated a normal distribution of attitudes, with responses being generally indicative of positive attitudes towards online assessment. Attia (2014), in a study of postgraduates' perceptions towards online assessment, reported that participants were, in general, quite satisfied with their online assessment experience.

2. Research Question

The overarching research question that the current study is exploring is: To what extent do candidates taking tests via an online-proctored mode feel satisfied with their experience?

3. Method

The main method used in the current study involves a survey administered to past candidates of LanguageCert's International English for Speakers of Other Languages (ESOL) suite of English language tests. This section reports the data collection procedure, with the survey administered via the Internet (and the *SurveyMonkey* online facility) from February 2021 onwards.

There are six tests in the IESOL suite, all of which are aligned to one of the six Common European Framework of Reference for Languages (CEFR) levels, A1 – C2. Due to language constraints, examinations offered in OLP mode are only available for candidates at B1 level and above.

The research team met in late 2020 and early 2021 to discuss issues related to the survey's design, with the survey worked on and revised during and after the meetings via email. The survey was then trialled on members of staff who had taken OLP examinations themselves. After moderating the survey and making modifications, the survey instrument was finalised.

Items are on a 6-point Likert scale, with '1' indicating a negative response or disagreement, and '6' a positive response or agreement. A 6-point scale was deliberately chosen to require respondents to commit to an opinion. The actual survey (see Appendix) consisted of 22 items in two sections. Section 1 (items 1-10) captured respondents' personal details. Section 2 (items 12-21) comprised 10 items in what was perceived as two broad sets. The first grouping could be loosely termed 'institutional', and centred around the delivery of the test by OLP means. The second grouping could be termed 'personal', and comprised items which probed respondents' views on issues such as anxiety, their experiences/reflections on the OLP process, and their preference for taking tests by traditional means or via OLP. Following a question soliciting assistance with follow-up structured interviews, a final open-ended question asked respondents for any additional comments that they might have on any aspect of the OLP process.

3.1 Ethical Considerations

The intention was to send an email link to the survey to all past LanguageCert candidates who had taken LanguageCert examinations via OLP, and who had agreed that they would be prepared to receive communications from LanguageCert.

In line with European Union General Data Protection Regulations, LanguageCert candidates state whether they are prepared to be contacted, or to receive any form of communication from LanguageCert, after taking an examination. Only candidates that agreed to being contacted were approached and were sent the link to participate in the survey.

4. Data and Analysis

The email link was sent to 7,170 candidates; it was opened by 2,917 and responded to by 920. The response rate of 31.5% was considered acceptable. Nulty (2008), in a summary of studies of both online and paper surveys, reports that online surveys in general achieve a rather lower response rate than paper-based surveys. He cites a figure, on average, of 33% for online, as against 56% for paper surveys. The current response rate of 31.5% comes very close to Nulty's figure.

A brief picture of respondent demographics will next be presented.

4.1 Demographics

This section presents a comparative picture of survey respondents versus the bigger picture of the entire cohort of LanguageCert candidates of IESOL tests. The IESOL test registration form asks candidates for details on gender, age, and mother tongue. Since not all candidates supply their details under normal circumstances, there is consequently a degree of missing data in the IESOL whole test figures. The survey, however, requested that respondents provide this demographic data, and all complied. Table 3 presents a comparison of respondents to the survey and typical pen-and-paper (P&P) candidate demographics taking B1 – C2 tests.

Table 3. Comparative picture of demographics in typical P&P test candidates and survey cohort

Demographics	Typical P&P IESOL Cohort (%)	Survey Cohort (%)
Test level		
B1	20.7	18.1
B2	38.3	29.8
C1	26.6	25.9
C2	14.3	26.3
Gender		
Female	53.5	63.9
Male	38.5	36.1
Age		
<21	35.1	10.1
21-30	36.0	27.0
31-40	16.6	28.4
41-50	7.9	24.3
>50	4.4	10.1
Mother tongue		
Greek	19.7	29.9
Italian	24.3	23.1
Chinese	13.9	7.0
Spanish	12.9	10.3
Polish		7.6
Other languages	29.2	22.1

Table 3 shows that the distribution of candidates across tests is broadly comparable to typical candidatures.

Comparatively more females than males take IESOL tests, a characteristic which is reflected in responses to the survey. We can also see that the survey cohort is older than the typical test population.

The mother tongue of approximately 70% of typical LanguageCert IESOL candidates is currently Italian, Greek, Chinese and Spanish. This pattern is broadly mirrored in the survey sample, although there are higher than typical response rates for speakers of Greek and Chinese. The higher response rate of the former is perhaps understandable, given that the survey was sent out from LanguageCert's Athens centre.

An analysis of the survey data will next be presented.

First, the robustness of the survey is gauged through reliability analysis and factor analysis. A presentation of key descriptives is then made – followed by an exploration of the inferential data. Finally, a section on qualitative data is presented. This comprises a thematic analysis of substantive open-ended comments provided by respondents.

4.2 Reliability

The ten attitudinal items on the survey achieved a reliability of 0.88 using Cronbach's alpha. Given that a level of 0.8 is generally recommended as desirable in a survey (e.g., Trobia, 2011), the survey's reliability was confirmed to be acceptable.

4.3 Factor Analysis

An exploratory factor analysis using Principal Component Analysis (PCA) with varimax rotation was conducted (working on the assumption that the underlying factors in the survey are related) to explore how the major constructs patterned out.

In line with Kaiser's (1974) recommendations regarding Sampling Adequacy Measures – the KMO (Kaiser-Meyer-Olkin) statistic – the figure of 0.91 indicated that the sample size was clearly adequate for factor

analysis. Results from Bartlett's test of sphericity – $\chi^2 (55) = 4,529.03$, $p < .000$ – indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. Only two components had eigenvalues over Kaiser's criterion of 1, which in combination explained 64.2% of the variance.

Taking loadings above 0.4 as being indicative of a cut-off point appropriate for interpretative purposes (Stevens, 2002), two possible factors emerge in the Component Matrix. Table 4 elaborates, with the items grouped together by factor.

Table 4. Component Matrix

<i>Components</i>	<i>1</i>	<i>2</i>
13. How clear were OLP setup instructions?	.839	
14. How straightforward was the OLP setup process?	.791	
15. How was the online connection with the interlocutor?	.722	
16. How clear was the interaction with the interlocutor?	.814	
17. How was the overall OLP experience?	.775	
18. Preference for tests by traditional means or by OLP?		.813
19. "Taking tests by OLP is a more personal experience"		.785
20. "Taking tests by OLP is more efficient"		.836
21. Your score: better on traditional or OLP tests?		.715
12. How anxious were you before the OLP test?		

As can be seen from Table 4, two clear factors emerged, with each factor, or latent trait, having at least four indicators (i.e., items), the minimum cut off recommended (see e.g., Yan, 2020). The factors that emerged may be defined as:

- (1) 'institutional' – comprising items 13, 14, 15, 16, 17, and centring around delivery of the test by OLP means.
- (2) 'personal' – comprising items 18, 19, 20, 21, and centring around personal reactions to taking a test by via OLP.

Item 12 – "anxiety" – appeared to be in a category of its own, not being included in either of the two factors identified. Such 'isolation' is mirrored in the responses below, as will be shown. The factor analysis would therefore appear to bear out the validity of the survey.

4.4 Descriptive Statistics

In the analysis below, responses of the 920 respondents in the sample are presented. Where possible, figures are matched against the general demographic trends of LanguageCert IESOL tests.

4.5 Attitudinal Items

This section is in two parts. First, items which diverge significantly from the mid-point of 3.5 are discussed. This relates to the issue of 'consumer validity' (Coniam, 2013), whereby a mean score considerably above (or below) the mean indicates strong acceptance (or rejection) of the proposition. In the current study, a '6' indicated a positive and '1' a negative response; in the current dataset (see Table 5 below), strong positive responses are defined as those above '4.5' while strong negative responses would be below '2.5' (although there are none of the latter in the dataset). Following the factor analysis, items are grouped into the factors identified.

Table 5. Survey item means (N=835)

<i>Survey item</i>	<i>Mean</i>	<i>SD</i>
12. How anxious were you before the OLP test?	3.3	1.7
06. Assessment of personal computer literacy 'institutional' items	4.9	1.1
13. How clear were OLP setup instructions?	5.1	1.3
14. How straightforward was the OLP setup process?	4.8	1.4
15. How was the online connection with the interlocutor?	4.7	1.5
16. How was the interaction with the interlocutor?	5.3	1.2
'personal' items		
17. How was the overall OLP experience?	5.1	1.2
18. Preference for tests by traditional means (1) or tests by OLP (6)?	4.8	1.6
19. "Taking tests by OLP is a more personal experience"	4.5	1.5
20. "Taking tests by OLP is more efficient"	4.7	1.4
21. Your score: better on traditional (1) or OLP (6) tests?	4.6	1.6

Item 12 had the lowest mean score, just below the central mean of 3.5. This is perhaps unsurprising, given that for many candidates, this was the first time they had ever taken an examination via OLP. Item 06 – a demographic question asking for an assessment of personal computer literacy – shows that candidates felt that they did not have problems working with computers or in interacting online. This suggests that the anxiety they felt may be attributed more to the looming examination than to how to respond via a computer. Nonetheless, the comparatively wider SD on item 12 (anxiety) is illustrative of the fact that, despite being computer literate, many are still concerned as they begin to take the examination.

Despite the anxiety many candidates clearly feel, the responses to the attitudinal items are all very positive. 4.5 has been proposed as a benchmark for endorsement of a proposition (Coniam, 2013). The positive nature of the responses may be seen by the fact that all the 'institutional' items have means in the high 4's or above 5. For the majority of respondents, the setup process was felt to be unproblematic; online connection was good; OLP setup instructions were clear; and interaction with the interlocutor was rated very highly indeed.

Responses to the 'personal' items were, on the whole, very positive, with the overall OLP experience in particular rated above 5. Respondents showed a clear preference for taking tests by OLP as opposed to by traditional means. Whether the extent of the preference for OLP has been spurred on by Covid-19, or is a sign of the times, remains to be seen. Nonetheless, the fact that taking tests by OLP was viewed to be a more personal experience may give an indication as to the greater uptake and longer-term acceptance of OLP.

Looking towards the future, on the issue of preference for tests by traditional means (1/6) or via OLP (6/6), a mean of 4.8/6 was recorded, indicative of very positive acceptance of OLP and strong future uptake of this means of test delivery. Respondents also felt that OLP was more personal and a more efficient way of taking a test.

4.6 Inferential Analysis

A chi square analysis of items where significant differences emerged will now be presented.

55 chi square analyses were conducted – the 11 attitudinal items against five background demographic variables. In general, little significance emerged on the majority of the analyses, indicating that respondents were in agreement with items irrespective of backgrounds such as gender, age, grade obtained, the level of test sat, or what test type – Speaking or Listening/Reading/Writing (LRW) test – had been taken.

There were only six instances of statistical significance, where 5% is taken as the level of significance. Table 6 elaborates, with the attitudinal item and variable affected identified.

Table 6. Attitudinal items – significant differences (sorted by item)

<i>Attitudinal items</i>	<i>Variables</i>	<i>Significance</i>
06. personal computer literacy	Gender (females)	$\chi^2(5) = 15.94, p = .007$
06. personal computer literacy	Age (>50)	$\chi^2(20) = 42.56, p = .002$
06. personal computer literacy	Level (B1)	$\chi^2(15) = 37.48, p = .001$
06. personal computer literacy	Grade (Fail)	$\chi^2(15) = 55.41, p < .001$
12. anxious before the OLP test	Grade (High pass)	$\chi^2(15) = 42.10, p < .001$
17. Evaluation of the OLP experience	Age (> 50)	$\chi^2(20) = 31.63, p = .047$

Of the six instances of statistical significance, four instances were related to perceived personal computer literacy – despite this item receiving an overall comparatively high mean of 4.9/6. Females and older respondents felt less technologically comfortable than men in general. Candidates who took a B1 test were in general less positive. As mentioned, the OLP process is conducted in English language, suggesting perhaps that for lower-level candidates – B1 in particular – the OLP experience puts extra demands on candidates with a lower ability in English. The fact that candidates who failed were more negative in their perceptions is perhaps understandable. It is possibly surprising that significance only emerged for fail grade candidates on one variable.

Significance was only reported against grade received, with, interestingly enough, high-pass candidates responding the most negatively.

Regarding reactions to the overall OLP experience, the oldest age group (above 50 years) was more negative than other age groups. Such response may reflect this age group's perception that they are also less computer literate than younger age groups.

4.7 Qualitative Data

In response to the final item requesting any comments from respondents, a total of 716 comments were received. After undergoing a thematic analysis, the comments were categorised, and totals were tabulated. Table 7 presents a picture of the results.

Table 7. Written comments provided (N=716)

<i>Total positive: 234</i>		<i>Total negative: 68</i>	
administration	9	administration	7
computer issues	6	computer issues	5
connection issues	6	connection issues	21
interlocutors	32	interlocutors	3
test materials	2	test materials	2
convenience	17	Covid issues	3
efficiency	7	test delivery	20
'minimalist'	157	other	7
total	234	total	68
"No comment"	394	Off topic	20

Of the 716 written comments received, 394 were framed simply as "no comment", "nothing to add" etc. 20 comments were classified as "topic", e.g., "A more accepted institute will be great, like Canada immigration." The latter two categories have been disregarded in the following discussion. 157 'minimalist' positive comments such as "very good", "thanks" were received; these have also been disregarded.

While the total number of positive comments outweighed the total number of negative comments by a considerable margin, in terms of *substantive* comments, the balance was about equal. In this context, O'Cathain & Thomas (2004) discuss how open-ended questions at the end of a survey help to "redress the power balance between researchers and participants". This is where respondents raise issues not covered by the closed questions – their "safety net" as Biemer et al. (2011) describe the open ended questions. In this light, the number of negative comments – as compared to the general tenor of positive responses to the closed questions – is understandable. This is the forum for negatively-oriented respondents to air any specific grievances they may have.

Table 8 presents a sample of comments on some of the key themes. Some themes received comments from both sides of the spectrum; some themes – connection issues, test delivery – were themes for negative comments.

Table 8. Major positive and negative Comment themes, with examples

<i>Key Theme</i>	<i>Positive tenor</i>	<i>Negative tenor</i>
Administration of exams	<ul style="list-style-type: none"> Just very happy with the service, they kept me informed about any changes, the people at the test centre very friendly and they followed the Covid strict standards. 	<ul style="list-style-type: none"> I didn't get my results in 3 business days as it is written in the home page of the website so I couldn't use this certificate for an exam and so I couldn't get a better mark in this examination. The most challenging and conflict part of my test was being told in the morning that my test has been cancelled on the short notice but also rewriting subject that I already passed.
Computer issues	<ul style="list-style-type: none"> My overall opinion on OLP tests is very positive, however I would like to report that my initial experience was a bit troublesome due to incompatibility of the program with Mac IOS. This should be made clear previously and maybe require the use of another system. 	<ul style="list-style-type: none"> Earphones should not be necessary.
Interlocutors	<ul style="list-style-type: none"> The lady who examined me was very nice and helpful. The internet connection wasn't very good, but thanks to her everything was alright. Thank you! 	<ul style="list-style-type: none"> I didn't expect proctors speaking with an inflection different from British or American one, so I struggled to understand her. I think you should solve the connection quality. The examiner was really nice and understood the situation but I felt really nervous cause there were a lot of breaks cause of the Internet. Anyway thanks a lot for this kind of certifications which allow us being prepared despite of Covid.
Connection issues		<ul style="list-style-type: none"> The internet connection in China mainland for online test is very poor. No matter I tried to use a VPN or not. And I have encountered several times that the application crashed. The examiner is kind and professional. I hope the examiner can consider the network factors of both sides in the oral test. If the examinee's performance is not good due to the network reasons, should we give a chance to retake the test?
Test delivery issues		<ul style="list-style-type: none"> At the start of my test, another candidate was quite close and I could hear a fair amount of background noise. This was not great. Listening part felt really loud, ExamShield wouldn't allow me to permanently lower the volume. Difficulties in the test: - the fact that we cannot print the test so that we can read it. It is not allowed to use a draft notebook.
Convenience	<ul style="list-style-type: none"> For me it was a great occasion to take this certification during a pandemic. I consider OLP exams a very efficient way to take exams, especially in this period of pandemic. 	
Efficiency	<ul style="list-style-type: none"> I would like to congratulate Language cert for the innovation way to examine students' English proficiency during Covid-19 lockdown. I also need to say a huge thank you to the personnel for the kindness, the patience and the excellent service. 	

Some of the issues raised above – such as scrolling, adjusting volume, seeing the how much of the test has yet to be completed is outlined in sample material and practice sessions on the LanguageCert website.

As mentioned above, it is in this comment section that the disaffected test takers may make their voice heard. The negative comments do nonetheless identify issues to be considered, and follow-up actions. One issue, which is under constant review, is that of internet connection. These have recently been upgraded and it is hoped that some of the issues raised by respondents here will have been addressed.

5. Conclusion

This paper has explored the reactions to and perceptions of candidates taking an examination via online proctoring, specifically in the context of English language examinations delivered by LanguageCert. A survey was sent out to LanguageCert candidates who had agreed to be contacted. 31.5% of candidates contacted responded to the request, in line with what might be expected for online surveys.

Demographically, responses were broadly comparable with typical LanguageCert candidates.

Responses to all attitudinal questions returned highly positive means, endorsing all aspects of the OLP process. Instructions, the setup process, interactions with the interlocutor – all received very positive responses. A relatively higher number of negative written comments were received compared to the generally positive tenor of responses to the closed questions. These have been passed on to the relevant PeopleCert departments. Some issues, such as internet connections, are constantly receiving attention from the systems section of the company.

Inferential analysis revealed computer literacy to be a significant concern for certain respondents. Females and older respondents appeared to feel less technologically comfortable – findings also reported in previous studies (see e.g., Yau & Cheng, 2012). B1-level candidates were less positive: the fact that the OLP process is conducted, and explained, in English may be an issue that needs to be considered. LanguageCert is currently looking at providing details to candidates in major languages other than English; this issue possibly needs to be given consideration regarding the OLP log-on, onboarding and setup process.

Regarding preference for tests by traditional means or via OLP, a strong endorsement of OLP was observed. Respondents felt that OLP was a more personal and efficient way of taking a test – probably as a result of the seamless continuity of test delivery via OLP throughout the Covid-19 pandemic. These positive signals clearly indicate a broad acceptance of OLP, pointing to strong future uptake of the OLP mode of test delivery.

There are nonetheless limitations to the current study, which future research might wish to address. One is that while the sample size in the current study was acceptable at 31.5%, the sample comprised past candidates who chose to open emails sent to them and who subsequently made the effort to open and complete the online survey. For some candidates, a considerable amount time had elapsed since they had taken the test. A further study should be conducted by candidates immediately post-exam. This would provide not only a larger sample, but would also catch immediate reactions and attitudes. It would also catch responses from more potentially disaffected candidates, who would not otherwise make the effort to complete a survey. A comparative study between the survey results presented in the current paper, and such a future study might shed interesting light on attitudes and opinions regarding online test delivery.

References

- Alexander, M. W., Bartlett, J. E., Truell, A. D., & Ouwenga, K. (2001). Testing in a computer technology course: An investigation of equivalency in performance between online and paper and pencil methods. *Journal of Career and Technical Education*, 18(1), 69-80. <https://doi.org/10.21061/jcte.v18i1.600>
- Ardid, M., Gómez-Tejedor, J. A., Meseguer-Dueñas, J. M., Riera, J., & Vidaurre, A. (2015). Online exams for blended assessment. Study of different application methodologies. *Computers & Education*, 81, 296-303. <https://doi.org/10.1016/j.compedu.2014.10.010>
- Attia, M. A. (2014). *Postgraduate students' perceptions toward online assessment: The case of the faculty of education, Umm Al-Qura university*. In Education for a Knowledge Society in Arabian Gulf Countries. Emerald Group Publishing Limited. <https://doi.org/10.1108/S1479-367920140000024015>
- Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., & Sudman, S. (2011). Measurement errors in surveys. *Technometrics*, 173(1). Hoboken. <https://doi.org/10.2307/1269316>
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, 54(2), 199-231. <https://doi.org/10.1111/j.1464-0597.2005.00205.x>

- Bonk, C. J., Lee, M. M., Reeves, T. C., & Reynolds, T. H. (Eds.). (2015). *MOOCs and open education around the world*. Routledge. <https://doi.org/10.4324/9781315751108>
- Cakrawati, L. M. (2017). Students' perceptions on the use of online learning platforms in EFL classroom. *ELT Tech: Journal of English Language Teaching and Technology*, 1(1), 22-30. <https://doi.org/10.17509/elt%20tech.v1i1.9428>
- Carnegie Council on Policy Studies in Higher Education. (1979). *Fair practices in higher education: Rights and responsibilities of students and their colleges in a period of intensified competition for enrollments*. Jossey-Bass.
- Clark, T. M., Callam, C. S., Paul, N. M., Stoltzfus, M. W., & Turner, D. (2020). Testing in the time of Covid-19: A sudden transition to unproctored online exams. *Journal of chemical education*, 97(9), 3413-3417. <https://doi.org/10.1021/acs.jchemed.0c00546>
- Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., & Thies, W. (2015). Detering cheating in online environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(6), 1-23. <https://doi.org/10.1145/2810239>
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214. <https://doi.org/10.1111/j.1467-8535.2008.00915.x>
- Eraslan, A., & Topkaya, E. Z. (2017). EFL students attitudes towards e-learning and effect of an online course on students success in English. *The Literacy Trek*, 3(2), 80-101.
- Foster, D., & Layman, H. (2013). *Online proctoring systems compared*. Webinar. Retrieved from <https://www.slideshare.net/caveonweb/caveon-webinar-series-online-proctoring-best-practicesoct-2013-slideshare-final>
- Gardner, L. (2020). *Covid-19 has forced higher ed to pivot to online learning. Here are 7 takeaways so far*. The Chronicle of Higher Education, 20. Retrieved from <https://www.chronicle.com/article/covid-19-has-forced-higher-ed-to-pivot-to-online-learning-here-are-7-takeaways-so-far/>
- Hos, R., Yagci, H., & Cinarbas, H. I. (2016). Turkish EFL students' perceptions about blended English courses in a teacher education program. *International Journal of Social Sciences and Education Research*, 2(3), 774-784. <https://doi.org/10.24289/ijsser.279022>
- Hussein, M. J., Yusuf, J., Deb, A. S., Fong, L., & Naidu, S. (2020). An evaluation of online proctoring tools. *Open Praxis*, 12(4), 509-525. <https://doi.org/10.5944/openpraxis.12.4.1113>
- Khan, R. A., & Jawaid, M. (2020). Technology enhanced assessment (TEA) in Covid-19 pandemic. *Pakistan Journal of Medical Sciences*, 36(19), 108-110. <https://doi.org/10.12669/pjms.36.COVID19-S4.2795>
- Lim, C. P., & Graham, C. R. (Eds.). (2021). *Blended Learning for Inclusive and Quality Higher Education in Asia*. Springer Nature. <https://doi.org/10.1007/978-981-33-4106-7>
- Lim, C. P., & Wang, L. (Eds.). (2016). *Blended learning for quality higher education: Selected case studies on implementation from Asia-Pacific*. Bangkok: UNESCO Bangkok Office.
- Mogey, N., Cowan, J., Paterson, J., & Purcell, M. (2012). Students' choices between typing and handwriting in examinations. *Active Learning in Higher Education*, 13(2), 117-128. <https://doi.org/10.1177/1469787412441297>
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & evaluation in higher education*, 33(3), 301-314. <https://doi.org/10.1080/02602930701293231>
- O'Cathain, A., & Thomas, K. J. (2004). "Any other comments?" Open questions on questionnaires—a bane or a bonus to research? *BMC medical research methodology*, 4(1), 1-7. <https://doi.org/10.1186/1471-2288-4-25>
- Özden, M. Y. (2005). Students' perceptions of online assessment: A case study. *International Journal of E-Learning & Distance Education/Revue internationale du e-learning et la formation à distance*, 19(2), 77-92. Retrieved from <https://files.eric.ed.gov/fulltext/EJ807820.pdf>
- Qualifications and Curriculum Authority. (2007). *Regulatory principles for e-assessment*. Retrieved from <https://publications.parliament.uk/pa/cm200607/cmselect/cmeduski/memo/test&ass/ucm3102paper4.pdf>

- Rahmawati, F. (2016). E-Learning implementation: Its opportunities and drawbacks perceived by EFL students. *Journal of Foreign Language Teaching and Learning*, 1(1), 47-53. <https://doi.org/10.18196/ftl.111>
- Rose, C. (2009). Virtual proctoring in distance education: An open-source solution. *American Journal of Business Education (AJBE)*, 2(2), 81-88. <https://doi.org/10.19030/ajbe.v2i2.4039>
- TPD@Scale for the Global South. (2020). *Teacher's guide for remote learning during school closures and beyond*. Information Technology Education and Development, Inc. Retrieved from <https://tpdatyscalecoalition.org>
- Watson, G., & Sottile, J. (2008). Cheating in the Digital Age: Do students cheat more in on-line courses? In *Society for Information Technology & Teacher Education International Conference* (pp. 798-803). Association for the Advancement of Computing in Education (AACE).
- Wiesenberg, F. P., & Stacey, E. (2008). Teaching philosophy: Moving from face-to-face to online classrooms. *Canadian journal of university continuing education*, 34(1). <https://doi.org/10.21225/D5JP4G>
- Williamson, B. (2015). Governing software: Networks, databases and algorithmic power in the digital governance of public education. *Learning, Media and Technology*, 40(1), 83-105. <https://doi.org/10.1080/17439884.2014.924527>
- Wright, J. C., & Kelly, R. (1974). Cheating: Student/faculty views and responsibilities. *Improving College and University Teaching*, 22(1), 31-34. <https://doi.org/10.1080/00193089.1974.10533496>
- Yau, H. K., & Cheng, A. L. F. (2012). Gender difference of confidence in using technology for learning. *Journal of Technology Studies*, 38(2), 74-79. <https://doi.org/10.21061/jots.v38i2.a.2>
- Zumbo, B. D. (2021). *A novel multimethod approach to investigate whether tests delivered at a test centre are concordant with those delivered remotely online: An investigation of the concordance of the CAEL [monograph]*. Vancouver, BC: Paragon Testing Enterprises/UBC Paragon Research Initiative, University of British Columbia.

Appendix

Online Proctored Tests: Experiences and Reflections

We would be very grateful if you could take a few minutes to reflect on the online proctored English language test that you took with LanguageCert. Please click on the circle, or select the number of stars as appropriate. **You do not need to identify yourself. All information collected is for research purposes only, and will be kept in the strictest confidence.**

OLP = online proctored; A 'Traditional Test' = a test by pen and paper; in a regular school or Test Centre setting; LRW = Listening, Reading and Writing

Section 1. Personal Details

01. I am	<input type="checkbox"/> Male <input type="checkbox"/> Female
02. I am years old	<21 21-30 31-40 41-50 >50
03. I live in (country)	
04. My mother tongue is ...	
05. My education level is ...	<input type="checkbox"/> Primary <input type="checkbox"/> Secondary <input type="checkbox"/> Bachelor Degree <input type="checkbox"/> Higher Degree
06. How computer literate do you consider yourself?	<input type="checkbox"/> not at all <input type="checkbox"/> very
07. The last OLP test I took was at level ...	<input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
08. How many LRW tests have you taken by OLP?	1 2 3 4 >4
09. How many Speaking Tests have you taken by OLP?	1 2 3 4 >4
10. What grade did you get on your last OLP test?	<input type="checkbox"/> Fail <input type="checkbox"/> Pass <input type="checkbox"/> High Pass <input type="checkbox"/> Prefer not to say

Section 2. Experiences and Reflections

11. Respond to the questions below EITHER (1) about Speaking; OR (2) about Listening, Reading and Writing (LRW)	I am responding about <input type="checkbox"/> Speaking <input type="checkbox"/> LRW
12. How anxious did you feel before your OLP test?	very anxious not anxious at all
13. How clear were the OLP setup instructions for the test?	not clear at all very clear
14. How straightforward was the OLP setup process?	very troublesome very straightforward
15. How was the online connection between you and the interlocutor during the test?	very poor very good
16. How clear were the interlocutor's instructions and directions during the test?	not clear at all very clear
17. How was the overall OLP experience?	very poor very good
18. Do you prefer to take tests by traditional means or by OLP?	prefer traditional prefer tests by OLP

19. "It is a more personal experience to take tests by OLP than to take tests by traditional means"	strongly disagree	strongly agree
20. "It is more efficient to take tests by OLP than to take tests by traditional means"	strongly agree	strongly disagree
21. Do you think that you score better in tests taken by traditional means or in tests by OLP?	better on traditional	better on OLP tests
22. Would you be available for a short follow-up (online) interview?	YES / NO. If yes, please leave your email or phone number.	
Do you have any comments that you would like to add?		

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).