'Time to figure out what to do': understanding the nature of Irish post-primary students' interactions with computer-based exams (CBEs) that use multimedia stimuli

Paula Lehane, Darina Scully & Michael O'Leary

Published online: 22 Feb 2022.

Submit your article to this journal ☑

Article views: 747

View related articles ☑

View Crossmark data ☑

Routledge
Taylor & Francis Group

'Time to figure out what to do': understanding the nature of Irish post-primary students' interactions with computer-based exams (CBEs) that use multimedia stimuli

Paula Lehane [ID], Darina Scully [ID] and Michael O'Leary [ID]

Dublin City University Institute of Education, School of Inclusive and Special Education, Dublin, Ireland

ABSTRACT

In line with the widespread proliferation of digital technology in everyday life, many countries are now beginning to use computer-based exams (CBEs) in their post-primary education systems. To ensure that these CBEs are delivered in a manner that preserves their fairness, validity, utility and credibility, several factors pertaining to their design and development will need to be considered. This research study investigated the extent to which the design of different types of test items (e.g. inclusion of multimedia stimuli) in a CBE can affect test-takers' engagement and behaviour. Qualitative data from a cued-Retrospective Think Aloud (c-RTA) protocol were gathered from 12 participants who had participated in a previous eye-tracking study. Participants watched a replay of their eye movements and were asked to state out loud what they were thinking at different points of the replay. Thematic analysis of the responses from these cognitive interviews captured the nature of students' interactions with online testing environments under three main themes: Familiarisation, Sense-making and Making Decisions. Students also provided their opinions of and recommendations for the future of Irish online assessments. These findings can offer guidelines to all stakeholders considering the use of CBEs in post-primary contexts.

Introduction

Consideration of how digital technology can be meaningfully integrated in a way that enhances learning has become a prominent topic in educational research literature in recent years (e.g. OECD 2015a; Ilomäki and Lakkala 2018), with an even more pronounced emphasis evident since the onset of the Covid-19 pandemic (e.g. Starkey et al. 2021). It is noteworthy, however, that a significant proportion of this research has focused on instructional practices (e.g. Cheung and Slavin 2013; Hoeffler and Leutner 2007; Tamim et al. 2011), with considerably less attention afforded to the design and use of assessments that incorporate digital technology.

CONTACT Paula Lehane   ✉ paula.lehane@dcu.ie

In the Republic of Ireland, examinations remain a dominant form of assessment at post-primary level. Until recently, all state examinations were administered in paper-and-pencil format. However, in May 2021, Ireland deployed its first computer-based examination (CBE) for its high-stakes Leaving Certificate Computer Science exam (SEC 2021). This reflects trends in other jurisdictions whereby CBEs are slowly being offered alongside, or indeed replacing, traditional paper-based exams (PBEs) (see, for example, NZQA 2018; 2021). Given the apparent satisfaction of key stakeholders with Ireland's pilot CBE (Donnolly 2021), and more generally, the increasing focus on digital technology in Irish schools in recent years (see DES 2015; DES 2017; DES 2020), it seems likely that CBEs will start to be deployed on a more widespread basis. If this is the case, much greater attention needs to be given to how such CBEs should be designed for this particular context. For example, there are many options in terms of the types of items that can be used in such tests, and the platforms that can be used to support them, but precisely how different choices in these domains might impact on the overall quality of the assessment remains poorly understood. Furthermore, little is known about Irish post-primary students' perceptions of CBEs as an assessment format, and consequently, their readiness for this transition. This seems like a missed opportunity, particularly in light of the well-documented potential of 'student-voice' to positively impact change (see Mitra, 2018).

With this in mind, this study sought:

(i) to better understand how particular types of test items can affect post-primary students' interactions with a CBE, from the perspectives of the students' themselves and
(ii) to gather the opinions and recommendations of these students in relation to the use of CBEs in the Irish education system

Computer-based exams (CBES)

CBEs are thought to have a number of advantages compared to their paper-based equivalents. Lehane (2019) noted that CBEs are considered more efficient than PBEs as they take less time to prepare, distribute and score. Others have noted that CBEs offer a more inclusive approach to assessment as digital environments offer a large range of accessibility features e.g. text-to-speech (e.g. Russell 2016). However, Lehane (2019, 6) also identified a number of threats to the use of CBEs for post-primary contexts including high development costs, inadequate school-based infrastructure, test security fears, and the time needed for stakeholders to adjust to this assessment approach. Concerns regarding technological failure are also prominent in the field (Cantillon et al. 2004). Regardless of these issues, there is a general consensus within the literature that CBEs are likely to become an inevitable part of the school lives of post-primary students worldwide.

Perhaps the key reason why enthusiasm for CBEs persists in the face of logistical concerns lies in the widespread belief in their potential to improve *validity*. Validity is regarded as the most fundamental consideration in developing and evaluating tests and assessments (AERA et al. 2014). Although the term was traditionally defined as 'the degree to which a test measures what it purports to measure' (Garrett 1937), recent definitions are more complex and holistic in nature, encompassing ethical considerations surrounding the consequences of test use (*e.g.* 'the meaningfulness and defensibility of the actions or decisions based on test scores, test-based information, or

assessment reports'; Chatterji 2013, 275). An understanding of this modern conceptualization of validity is particularly important when considering high-stakes examinations such as the Irish Leaving Certificate.

Evidence of validity can be gathered from many different sources. *Construct validity evidence* can be obtained by considering the authenticity with which the target knowledge, skills and abilities (collectively known as the test 'construct') are captured. Russell (2016, 21) asserted that the purpose of any question (or 'item') in an assessment is to 'collect evidence of the test-taker's development' of the target construct, and while PBEs are often limited to short-answer, essay-type or multiple choice questions, CBEs, in contrast, can include a greater variety of multimedia stimuli. 'Multimedia' refers to the combination of text with other media elements such as images, animations or simulations to communicate meaning and information (Jordan 1998). The use of multimedia stimuli has allowed researchers to assert that CBEs can present more 'authentic' contexts for test-takers to demonstrate their knowledge, skills and abilities (e.g. Parshall and Harmes 2008). However, while there is a 'broad faith' amongst educationalist that CBEs are preferable due to this property, the exact nature of this value, if it even exists, is difficult to verify and describe (Bryant 2017, 1) given the limited research in the field. The available research on the use of multimedia stimuli in assessment contexts, and its potential impact on validity, will now be interrogated.

Using multimedia stimuli in CBEs

Literature suggests that the addition of multimedia stimuli to a test item may affect *construct measurement* as well as test-taker behaviour and performance. Work by Lindner et al. (2017) indicated that the addition of representational pictures (illustrations that visualised the item stem without adding solution-relevant information) to text-based items in an online test of scientific literacy improved student performance, accelerated item processing and reduced rapid guessing behaviours. While the authors did suggest that this facilitative effect was a positive one for assessments, they also conceded that the inclusion of images in the test items may have 'taken away the need [for test-takers] to build mental visualisations' (Lindner et al. 2017, 491) thus changing the constructs being assessed. Other work by Lindner and colleagues (2021) has also shown that the inclusion of pictures in test items can influence test-takers' thoughts and opinions about their expected achievement. These findings suggest that the inclusion of the most basic multimedia objects in an item can modify the behaviours that test-takers engage in which may have significant implications for validity.

Images are not the only form of multimedia objects available to CBEs. Animations can also be included. Animations depict a 'simulated motion picture … [showing] movement of a drawn (or simulated) object' (Mayer and Moreno 2002, 88). Animations can present a more realistic picture of a given situation and can often communicate complex concepts and information more efficiently than text (Tuzinski 2013). When replacing text, animations can also reduce the likelihood that a test-taker's ability to process text will have an undue influence on their performance in a test that is designed to measure something else entirely. This phenomenon is known as 'construct-irrelevant variance' and represents a serious threat to validity (Messick 1994). Karakolidis et al. (2021) compared the performance of native ($n = 51$) and non-native ($n = 66$) English speakers taking an

animated and text-based version of a situational judgement test measuring teachers' interpersonal skills. The variance attributed to construct-irrelevant factors like native language and reading comprehension was lower by 9.4% in the animated version. However, a clear understanding of how animations can change the nature of test-takers' interactions with a test item has yet to be obtained.

A third variation of multimedia objects is now becoming more commonplace in CBEs – simulations. Simulations are interactive forms of multimedia objects whereby test-takers can 'produce' an imitation of a real world scenario (Levy 2012). Simulations can assess a test taker's proficiency on the basis of their interactions with the virtual environment and their ability to use information they have generated to answer other items (Baker and Clarke-Midura 2013). Consequently, simulations hold a significant amount of potential for the assessment of some of the more complex cognitive processes of Bloom's taxonomy, specifically 'analysis' and 'evaluation' which can be difficult to achieve using traditional exam questions (see Scully 2017). While use of these items is growing in popularity (e.g. OECD, 2017), some commentators have noted that their introduction to educational CBEs has been somewhat rushed from a practical and psychometric perspective which poses a threat to the validity of inferences drawn (e.g. Shiel et al. 2016; Lee et al. 2019). In particular, there appears to be a lack of understanding as to how test-takers engage with these multimedia objects (Teig et al. 2020).

Test-takers' views on and readiness for CBEs

As previously argued, construct-irrelevant variance is a major threat to test validity. Messick (1994) noted that the testing process may contribute to construct irrelevant variance, especially if test-takers' unfamiliarity with a test's *format or procedures* significantly affects their ability to engage in the processes necessary to demonstrate their competence. Huff and Sireci (2001) outlined a number of ways that the process involved in the administration of CBE could contribute to construct irrelevant variance e.g. inadequate computer proficiency, inadequate computer platform familiarity, anxiety due to a change in test format. Indeed, test-takers themselves also seem to be aware of the potential risk of CBEs to unfairly impact their performance. For example, studies conducted in higher education contexts have shown that students can be reluctant to take a high-stakes CBE as they fear that the change in administration format would prohibit them from using their preferred test-taking strategies (e.g. making notes, skipping questions; Hochlehnert et al. 2011). However, research by Deutsch et al. (2012) demonstrated that third-level medical students in Germany were shown to have a more positive attitude towards CBEs after taking practice exams and items. Practice exams allowed test-takers to become familiar with the CBE's User Interface (UI) and the type of questions that could be asked. Therefore, the value of test-taker familiarity and comfort with this administration mode needs to be considered in order to safeguard validity.

User feedback to support familiarity and comfort with CBEs can aid in the development of CBEs that support valid inferences. For example, undergraduate students participating in Walker and Handley's (2016) research considered easy navigation to be essential to the usability of an online examination. Free movement of students in a CBE was deemed necessary to accommodate students' test-taking strategies. Taking into account such preferences may ensure that test-takers are more positively disposed

towards the introduction of CBEs. However, very few studies on the views of post-primary aged test-takers appear to exist. This is surprising given the current and expected scale of CBE use for this age group. The views and preferences of third-level students appears to dominate the field. While the NZQA (2018) did gather a large amount of user feedback from their post-primary test-takers (which then subsequently informed future iterations of their CBE), other peer-reviewed studies are more difficult to find. While some tangentially related research does exist (e.g. Siozos et al. 2009), far more is needed.

This study

The current study was a part of a larger piece of research that set out to explore the design of test items in CBEs and how test-takers' perceptions of and interactions with items of varying designs differed. Specifically, the research questions for this study were:

1. What thought processes underlie post-primary test-takers' interactions with static, dynamic and interactive multimedia items in a CBE?
2. What recommendations for future CBEs can Irish post-primary test-takers provide to relevant stakeholders?

Methods

Design

As the purpose of this study was to gain an in-depth understanding of the processes underlying students' interactions with CBEs, as opposed to focusing on their performance on the exam, a qualitative research design was deemed most suitable. All participants completed a CBE and were then invited to engage in a cognitive think-aloud.

Measure and procedures

CBE for scientific literacy

Publicly available assessment units from the domain of scientific literacy within the Programme for International Student Assessment (PISA) were used to create a CBE (OECD 2015b; 2017). These assessment units consist of stimulus materials (e.g. text, tables, diagrams etc.) followed by one or more items based on the stimulus materials. Each available unit was based on an applied area of scientific knowledge (e.g. ecology) and the items contained therein availed of a variety of response actions e.g. multiple choice questions, simulation-type items. The items in these units aimed to assess the general scientific literacy skills that students aged between 14 and 16 years are expected to have. Units 1–5 made up Part A of the CBE and Unit 6 was Part B of the CBE.

1. Bird Migration (1 practice item)
2. Meteoroids and Craters (4 items)
3. Sustainable Fish Farming (3 items)
4. Blue Power Plant (4 items)

5. Groundwater Extraction and Earthquakes (4 items)
6. Running in Hot Weather (5 items)

   Participants completed a CBE that contained items with either static (images) or dynamic (animations) multimedia stimuli (Part A[1]), followed by five simulation-type items (Unit 6; Part B). For the dynamic condition in Part A, each animation displayed a moving representation of the key concepts and ideas as a voiceover read aloud the text of the original item. To ensure the quality and accuracy of the animations, the researcher and three different people with expertise in educational research and/or post-primary science content reviewed each of the animations. These units were contained within a bespoke CBE platform as off-the-shelf commercial platforms did not meet the necessary requirements. The CBE's testing platform for Part A was similar to the PISA 2015 platform in terms of appearance and functionality. Figure 1 shows the platform used by the current study for one item for both conditions. For Part B
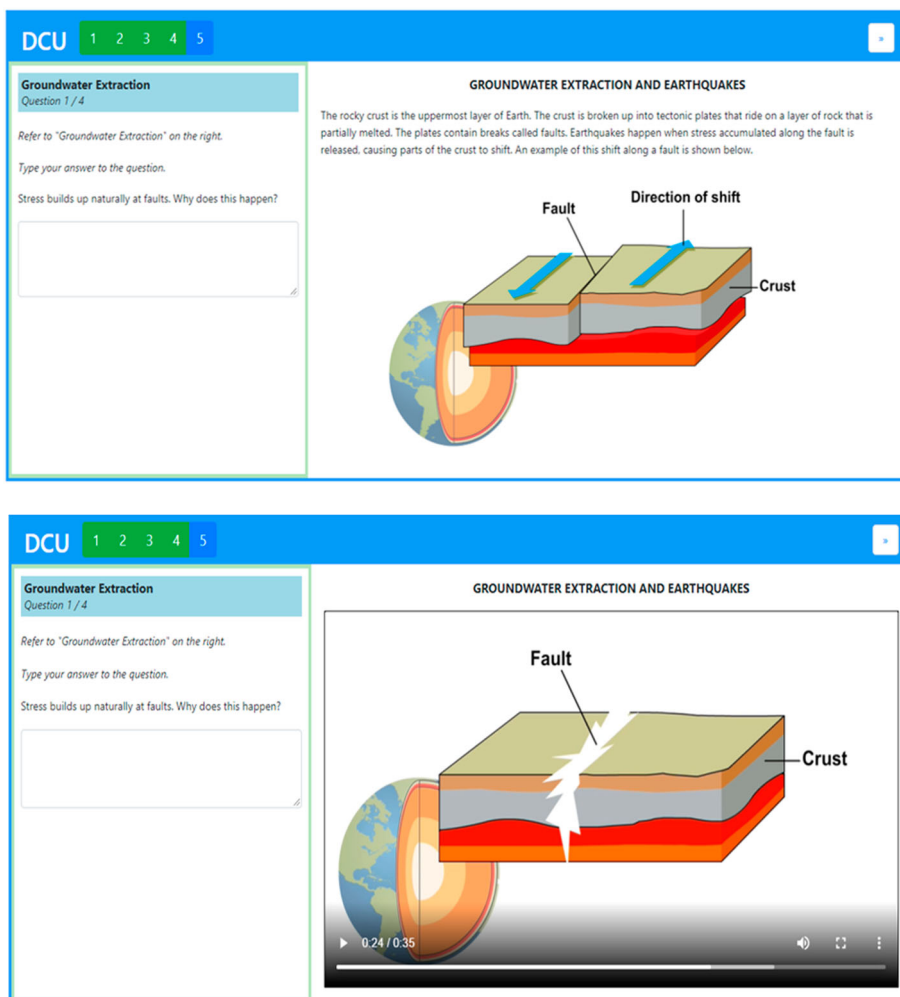

Figure 1. Item from 'Groundwater Extraction and Earthquakes' unit (static, dynamic).

('Running in Hot Weather'), the interactive items in this unit were accessed by participants directly on the OECD website as these could not be hosted within the bespoke CBE platform (see http://www.oecd.org/pisa/pisa-2015-sciencetest-questions.htm).

Eye-Movement replay

Whilst completing the CBE, participants' eye movements were monitored by the Tobii Pro Fusion eye-tracker (120 Hz). This provided a record of participants' eye movements in the form of fixations (the stable state of the eye) and saccades (movements between fixations). Fixations, generally, represent the focus of an individual's attention and saccades show the change in the focus of visual attention (Alemdag and Cagiltay 2018). The eye-tracker provided a gaze plot video for each participant outlining their location and order of the fixations and saccades overlaid on each test item.

Cued-retrospective think aloud

Following completion of the CBE, participants were presented with a gaze plot that replayed their eye movements whilst simultaneously engaging in a think-aloud interview. Think-aloud methods ask participants to verbalise their thoughts on a task, which can be used to better understand the mental processes that underlie an individual's performance (Salkind 2010). These approaches have been used extensively in cognitive science research and have been particularly useful in explicating expert and novice performances across a range of tasks e.g. chess (Salkind 2010). As completion of a CBE involves engaging in a number of complex tasks, a think-aloud protocol was considered a suitable qualitative data collection tool. Taking into consideration that cognitive processes are quicker than verbal processes (whereby participants may be thinking about more than they can verbally express) and that the act of trying to verbalise thoughts may also interfere with task performance (Olsen et al. 2010), a *retrospective* think-aloud (RTA) approach was considered the most appropriate. RTAs require participants to remember their experiences rather than communicate their moment-to-moment decisions and actions as they happen. However, such an approach does mean that important information may be forgotten or misremembered (Elbabour et al. 2017). To counteract this, a specific type of retrospective think-aloud was deployed in this study - a cued-RTA (c-RTA).

In this study the 'cue' was the eye movement video. Each participant was presented with a replay of their eye movements and asked to recall their thoughts and actions for four items (one multiple-choice item, one item requiring a drag-and-drop response, one open-ended text response item and one simulation-type item).

Participants

Twelve Transition Year students from a rural post-primary school in Ireland participated in this study. The average age of the participants was 15.6 years (SD: 0.6). In relation to their performance on Part A of the CBE, four participants achieved a 'low' score (<55%), seven achieved a 'moderate' score (>55%) and one participant was classified as a 'high' scorer (>75%). In relation to the simulation-type items (Part B), there was one participant with a 'low' score, five with 'moderate' scores and six participants attained a 'high' score (see Table 1).

Table 1. Profile of participants.

	Condition	CBE Part A	CBE Part B
Participant 1 (P1)	Static	73% (M)	81% (H)
Participant 2 (P2)	Dynamic	40% (L)	6% (L)
Participant 3 (P3)	Static	100% (H)	81% (H)
Participant 4 (P4)	Dynamic	67% (M)	63% (M)
Participant 5 (P5)	Dynamic	67% (M)	63% (M)
Participant 6 (P6)	Dynamic	67% (M)	81% (H)
Participant 7 (P7)	Dynamic	53% (L)	81% (H)
Participant 8 (P8)	Dynamic	67% (M)	100% (H)
Participant 9 (P9)	Static	60% (M)	69% (M)
Participant 10 (P10)	Dynamic	53% (L)	63% (M)
Participant 11 (P11)	Static	47% (L)	56% (M)
Participant 12 (P12)	Static	73% (M)	88% (H)

Data analysis

The data collected from these participants were analysed using Braun and Clarke's (2006) six-step framework for thematic analysis. Table 2 outlines how each step of Braun and Clark's (2006) framework was applied. NVivo 12 software (QSR International 2020) was utilised to facilitate the process.

Findings

Figure 2 summarises the final thematic framework that was constructed based on the qualitative data. Both latent and semantic themes are present in the thematic framework as per Braun and Clarke's (2006) definitions. The three latent themes identified ('*Familiarisation*', '*Sense-making*' and '*Making Decisions*') 'examine the underlying ideas, assumptions and conceptualisations' within the data (Braun and Clarke 2006, 84). These themes explain the nature of test-takers' interactions with the CBE. The fourth theme ('Feedback') is a semantic theme whereby 'the analyst is not looking for anything beyond what a participant has said' (Braun and Clarke 2006, 84) and instead seeks to show patterns in content so as to highlight the significance of the patterns and their broader meanings and implications' (Braun and Clark 2006, 84).

Theme 1: Familiarisation

The first latent theme, *Familiarisation*, reflects the means through which participants orientated themselves to the online testing environment and the overall

Table 2. Application of Braun and Clarke's (2006) six-step framework for thematic analysis.

| Step	Description/Actions
Familiarisation	Audio recordings (233 min) were transcribed. These transcripts were read multiple times by the first author to facilitate familiarisation.
Initial Coding	The data were then organised into smaller 'chunks' or 'codes' in a meaningful and systematic way using an inductive approach.
Theme Search	The initial codes were then organised into broader themes. These were descriptive in nature.
Theme Review	The initial themes were modified and refined and the data contained within each theme were reviewed.
Theme Definition	Efforts were made to ' … identify the essence of what each theme is about' (Braun & Clarke, 2006, 92). Relationships between and within themes were also identified.
Write-up	The final thematic framework was finalised after discussion with all authors.
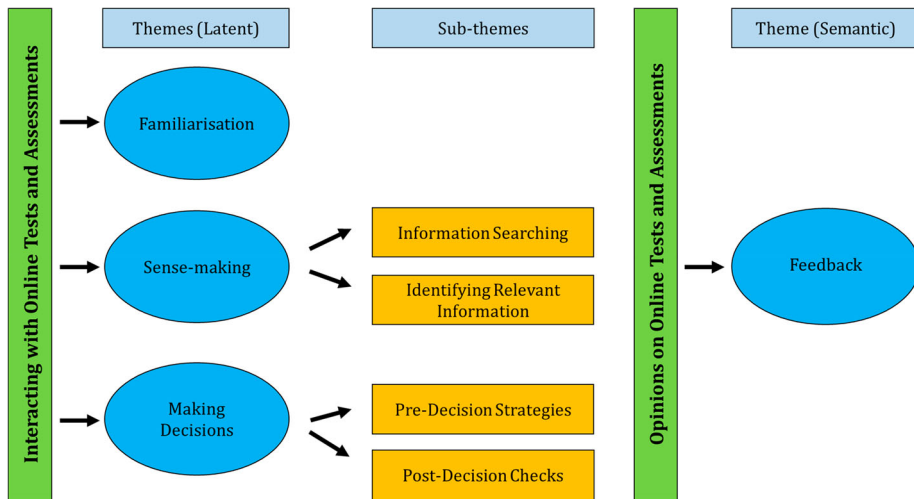
Figure 2. Thematic frame representing principal themes and subthemes.

value they placed on this process. All of the participants mentioned that it was important to first have '*time to figure out what to do*' (P2). During this time, participants asserted that any confusion regarding the overall layout of the system needed to be overcome as soon as possible e.g. how to select an answer. Participants noted that the volume of information on the screens during these practice items required them to actively pause and search for the spatial *position* of key elements (e.g. question, response options) to use as 'checkpoints' *before* cognitively engaging with them.

> I was just looking all around the screen cos I was trying to find that actual question. **P4**

> I took a minute to get used to it 'cos there's a lot on screen. **P8**

All of the participants noted that the practice items were necessary and valuable. Yet, the familiarisation process appeared to be repeated each time a *new* item type was encountered. For example, when reviewing their eye movements for the 'drag-and-drop' item, some participants admitted that they ignored the instructions explaining how the relevant objects should be moved and ordered as they did not have a similar appearance to the instructions found in other items (i.e. they were not italicised). Instead, they immediately engaged with the test stimuli to understand how the question could be completed before engaging with the actual content of the test item. Others noted that their understanding of drag-and-drop type items was aided by their experiences with other online environments e.g. dragging and dropping browser tabs (P9), playing games on the DS (gaming console; P7). One participant realised that they failed to transfer their knowledge from non-assessment online platforms (P2). Regardless of their prior knowledge of online environments, becoming more familiar with the testing platform and the test items was appreciated by the participants as it allowed them to develop more efficient information search strategies.

> Um, yeah. But after the third question, I kind of basically knew exactly what to do and where things would pop up. **P2**

Like, initially I thought, like, 'oh wow, that's a lot of things on the screen'. But by this question [QUESTION 3], it was more manageable then. **P9**

Theme 2: Sense-Making

The second latent theme, *Sense-Making*, captures the thoughts and behaviours participants engaged in when attempting to sort and use the information presented to them. Two distinct approaches to the sense-making process were identified: *Information Gathering* and *Identifying Relevant Information*.

Information gathering

When the participants encountered a test item, they engaged, for a relatively brief period of time at least, in a general visual search. This was different from what was described in the *Familiarisation* process as the participants now took into consideration the *content* of these elements. The visual stimuli (videos, text and images) acted as an important reference point to guide participants in their efforts to understand the test item's content. However, different searching techniques were employed depending on the type of visual stimuli presented. For the participants in the dynamic condition, the videos were generally ignored after they had been played once. The video was always played for the first item and then occasionally in later items if participants wished to double check something. Furthermore, the visual elements of the videos were largely ignored when they were first played. Instead, the seven participants in the dynamic condition listened to the audio narration while they read the test item and/or response options. Some participants in the dynamic condition were aware that this use of the audio narration was a key aspect in their information search strategy with one noting that ' ... *the first time I played it, I was mainly listening to the video'* (P6). Others only became aware of this behaviour when they reviewed their eye movements.

I'm actually reading instead of looking at the video. Didn't realise that. **P5**

Those in the static condition acknowledged that the way in which information was presented to them on-screen was similar to that of a 'paper-and-pencil' test (e.g. P3, P9). Interestingly, some participants in the static condition felt that the images accompanying the text were not always useful. Some noted that the images in the *Power Plant* and *Groundwater Extraction* units were the *only* beneficial ones as they were diagrams rather than pictures. Diagrams were highly valued by the participants in the static condition as they gave '*the gist of what you're going to be working on'* (e.g. P9). Indeed, as indicated by more than half of the participants in the dynamic condition, static diagrams were so desirable in the CBE that these participants often created their own diagram by pausing the animation. P8 explained that pausing the videos in the dynamic condition (and thus, unintentionally creating a static stimulus) was a more efficient way for them to gather information when the videos were explaining diagrams.

Identifying relevant information

In general, once participants had gained an understanding of the overall position and content of the test item's elements, the process of identifying *relevant* information (i.e. any information that they believed would help them to complete the test item) began. Some appeared to have 'transferred' over from paper tests and others were specific to the testing environment

or condition. Unsurprisingly, most participants admitted that they spent some time trying to recall what '*they already knew from Science class*' about a particular topic (P7). Furthermore, those in the static condition attempted to find and match key words from the stimulus text and the question stem. Participants in this condition noted that this was a standard test-taking strategy that they felt comfortable using in an online environment. They did this consistently in every item, even for those items where the stimulus text was the same as the previous item.

> I'm just trying to find a keyword that I can find in both the text and answer. I always do that in every question. **P9**

For those in the dynamic condition however, there was no opportunity to do this. Instead, these participants had to 'listen out' for the keywords in the audio narration. To speed up this process, all of the seven participants interviewed noted that they skipped through the video listening out for a key word or visual cue. For some, the use of videos as a presentation format was frustrating as '*you don't get the information immediately*' (P8). Others felt justified in not playing the video more than once in a unit as they '*didn't need all of it again*' (P8) or '*remembered the content pretty well from the last two questions*' (P10). Although the videos often slowed down their search for relevant information, they did appear to provide other contextual information that students considered relevant. In attempting to determine which energy conversions occurred in the *Power Plant* unit, P5 said that they '*remembered the lights lighting up*' in the video, thus making them more confident that at least one of the energy forms involved was electrical. Similarly, P8 highlighted that skipping to the end of the video allowed them to watch how '*the electricity came on after the water made the turbine move*'. This gave them the information they needed to complete one of the test items. This animated representation of energy conversions in the *Power Plant* unit may have allowed participants in the dynamic condition to more easily identify the relevant types of energy involved compared to those in the static condition.

For the simulation-type items, some participants admitted to identifying in advance what areas of the simulation output they should attend to *before* running the simulation (e.g. P4, P5, P9). However, not all participants employed such a focussed approach, with others waiting until *after* the simulation had been run to look for the relevant information needed (e.g. P1, P6, P11). Participants had to sort through a large amount of information to identify the relevant information they needed to complete the simulation-type test. In attempting to identify this information, personal preferences seemed to play some role.

> I just looked at the images on the top … and then I tried to remember which one was right … I barely ever looked at the table. **P5**

> I thought the bits up top were a bit useless. The table was more useful in deciding the answer cos you had a record. **P7**

Theme 3: Making decisions

The third latent theme, *Making Decisions*, represents the decision making process undertaken by the participants as they completed each test item. Two key stages to this process

were recognised: *Pre-Decision Strategies* and *Post-Decision Checks*. The first of these embodies *how* the participants came to their final decision based on the information they had previously deemed relevant. The second represents the final interactions the participants had with an item before moving onto the next one.

Pre-decision strategies

When reviewing their eye movements, participants recalled their thoughts in selecting or constructing their final response to an item with relative ease. In making a final decision on an answer for a test item, some participants did admit to guessing if they were unsure (e.g. P2). However, for multi-part questions, such as those that needed participants to select two words to complete a sentence, this uncertainty was much easier to manage. If participants knew the answer to one part of the item, they answered that part first and then considered the other part of the item. The participants acknowledged that this approach of '*start(ing) with the answer you are more confident of*' (P7) was one they would employ in a standard pen-and-paper exam. However, it was much easier to use this strategy in an online exam.

> It's just two clicks. It doesn't … It's kind of quicker than just rubbing something out and stuff. It's no big deal if you change your answer or just put down a placeholder in an online exam. **P9**

Other strategies to support their final decision were also described by the participants. For multiple-choice questions, the participants often 'eliminated' the possible response options one-by-one, even when they were confident of their answer. P8 admitted that they knew immediately that three of the options could be eliminated but they '*needed to read it twice to make sure*'. This preference for 'double checking' information before making a final decision was evident regardless of item type.

Post-decision checks

Deciding upon a particular answer or response option did not signify the completion of a test item. Analysis of the qualitative data indicated that the time *after* making a decision on their final response to an item but *before* moving onto a new item was distinguished by a number of key behaviours among the participants. The participants reported that they spent some time checking the item one last time before moving onto the next test item. The behaviours associated with these 'post-decision checks' were very similar to those that constituted the pre-decision strategies. For example, after completing an item many participants spent some time '*double checking*' their answers one last time (e.g. P1, P11). This occurred even when the participant had been confident in their final decision. When queried further on this, one participant noted that they would '*always do this in a test*' (P12) and were just transferring previously taught test-taking strategies to the online environment. However, at least half of the interviewed participants indicated that this interaction with a test item was a new experience for them that was prompted by the online environment.

> **P3:** Like, if I was in an exam, I usually just go over something [*sic*] once at the end so that I have enough time during the exam. I wouldn't like double check it straight away.

Researcher: So were you more likely to double check it on the computer?

P3: Yeah, that's the reason I did good here I think.

The testing system did not allow the participants to review their answers before submission. While the participants were aware of this from the outset, this did not seem to be a factor that contributed to the occurrence of these post-decision checks as no participant mentioned it. Furthermore, it was observed that the participants rarely changed their answers during the post-decision time period, thus suggesting that it was not likely that uncertainty over their answer prompted them to double check their work again. Instead, the online environment itself seems to have naturally encouraged the participants to do some post-decision checks. According to the participants, online testing environments were considered more legible than traditional paper-and-pencil approaches. Interviewee 11 noted that '*it's easier to see and spot stuff online than written down in your own handwriting I think*'. This opinion was supported by other participants too (e.g. P10).

Theme 4: Feedback on CBEs

The final theme addresses participants' feedback on the online test they had just completed and their view on CBEs in general. Participants offered a number of recommendations for the design and use of online tests during and at the end of the CBE. For example, participants had a clear preference for online exams compared to pencil-and-paper tests. However, there appeared to be some conditions attached to this preference. Participants were predominately in favour of online tests for subjects that required them to generate a large amount of text. Online tests would allow them to type instead of handwriting the answers. This was preferable as typed text was considered to be '*neater, quicker and easier*' (P10). However, at least five of the interviewed participants recognised that their own typing skills would need to be addressed before they would be comfortable with CBEs.

> I'd need to learn to type properly to be happy with an online exam for the Leaving [Certificate]. **P12**

Many participants recommended that high-stakes exams for some subjects be excluded from online platforms e.g. geography, engineering, mathematics. Most of the participants indicated that the activities required of them in an exam for these subjects are difficult to do on a screen e.g. drawing diagrams, writing formulae etc. As a result, they recommended that online exams for these subjects not be considered.

> Um, maths ... it's just really practical and you have to write formulas down. ..And in geography, you have to draw loads of diagrams. **P7**

> English or history would be OK to do because they have a lot of typing. **P10**

> Uhm ... maybe not woodwork? Because you have to do some sketching. **P11**

In relation to the actual design of CBEs, the participants did provide some interesting insights. For example, it appeared that there was no real preference for one item type over another. In fact, one participant noted that they '*liked the variety*' (P10). Interestingly, two participants from the dynamic condition noted that they would have preferred to have seen text-based stimuli rather than the audio-visual stimuli they had experienced. P4 felt that for the majority of the videos '*the picture was enough at the end*'. P2 argued that the absence of text to refer to made some of the items '*really hard*'. Other participants did not note anything of significance in relation to the use of video-based or text-image stimuli in the test. In contrast, participants did make an effort to note that, regardless of an item's type or its content, careful consideration should be afforded to how an item looks on a screen. Half of the participants recommended that certain aesthetics should be adhered to when designing an online screen to make is easier to interact with the test platform. These design recommendations usually related to the use of '*specific font types to indicate different things*' (P5). P3 recommended that '*questions should be in a different font and bold so that you can tell what's a question and what's just random*'. Another participant suggested that having blank sections and spacing between elements is important to prevent students from feeling '*overwhelmed*' (P8).

Other general recommendations for the overall design of an online test were also highlighted. Two participants said that they felt reassured by the system's warnings if they had not answered a question properly or forgotten something. Yet, despite this, the participants did suggest that more navigational freedom e.g. being able to skip questions and then return to them, was needed in online exams, particularly in comparison to the test they had just completed.

> You don't really have an option of skipping anything online but you might want to do some parts first. You need to be able to skip to them. **P12**

> I knew where everything was and normal tests ... it's happened before where I missed an entire page! **P5**

Discussion

This study examined the thought processes that underlie test-takers' interactions with a CBE that used a range of multimedia stimuli. It also elicited the views of post-primary test-takers in relation to this type of exam.

Three latent themes captured the nature of test-takers' interactions with technology-based test items. Each theme provided a richer and more detailed understanding of how post-primary test-takers engage with test items in online environments. In the 'Familiarisation' theme, test-takers acknowledged their increasing comfort and fluency with test items with repeated exposure. Familiarity with test items appeared necessary to support effective test-taking strategies and to ensure test-taker comfort with CBEs. By having some familiarity with the content and layout of the testing environment in advance of using it in a high-stakes context, test-taker ease with the CBE is likely to be

increased. Consequently, construct irrelevant variance caused by test-taker anxiety or uncertainty is likely to be reduced (Wise 2019). Therefore, to safeguard a CBE's validity, familiarisation activities should be provided in advance of a high stakes CBE. This aligns with current wisdom on the subject (NZQA 2018) and reflects the findings of research from third-level students (e.g. Deutsch et al. 2012).

The 'Sense-making' theme provided similar insights. Sense-making is a concept originally derived from organisational theory (Weick 1995) but has been applied in many different contexts to describe the actions preceding a judgement or decision. Interviewees revealed that they could 'make sense' of an item in the dynamic condition by gathering information from the item's stimulus through their auditory channel while simultaneously obtaining information on the contents of the interaction space through their visual channel. This aligns with Mayer's (2008; 2014) Cognitive Theory of Multimedia Learning. However, the data also indicated that test-takers experienced different barriers in their search for relevant information when in the dynamic condition e.g. not being able to 'find' key words. Test-takers also had preferences for different visualisations e.g. static diagrams. These individual differences in the sense-making process supports the idea that the design of items, including the use of different multimedia objects, in CBEs can affect test-takers interactions with a CBE, which may affect validity. This finding justifies previous concerns on the introduction of construct irrelevant variance to testing context as a result of multimedia stimuli (e.g. Bryant 2017; Huff and Sireci 2001).

Insights into individuals' decision-making process were also obtained as a result of the qualitative data gathered. These should offer some comfort to those considering the introduction of CBEs to post-primary settings. Test-takers revealed that, wherever possible, they would 'transfer' the strategies that they would use in a standard pencil-and-paper exam to a digital test item e.g. answering multi-part questions out of order etc. The test-takers involved in this study did not appear to have much difficulty in transferring the majority of their test-taking strategies to this online platform. Furthermore, they noted that 'post-decision' behaviours were easier to execute in an online environment. For example, 'double checking' decisions rarely resulted in an answer change but participants highlighted how easy it was to do in the online environment. This particular interaction with CBEs had not been previously highlighted by research.

A number of specific recommendations regarding design and deployment of CBEs for Irish post-primary students were identified. Test-takers in this study believed that a sound interface design was essential for success in CBEs. Harms and Adams (2008) asserted that each component of an online interface must be designed 'with consideration of the knowledge, expectations, information requirements, and cognitive capabilities of all possible end users' (p. 4). Therefore, the interface of a CBE should take into consideration the specific needs of students in an online testing environment. The test-takers in this research highlighted some of these needs, including 'warnings' if a question had been forgotten, the freedom to navigate between items and the ability to review their final responses. Other pre-requisites for the use of CBEs were also identified e.g. typing proficiency, content suitability. Such information on the needs and preferences of post-primary aged test-takers in relation to CBEs has not been reported in literature thus

highlighting the contribution of the current research to the field and its potential value for those involved in the design and deployment of post-primary CBEs.

Limitations and future research

This research examined one form of digital assessment – CBEs. However, it is important to acknowledge that there is a growing appetite for broader range of formative and summative digital assessment tools to be used at post-primary level (Berry 2011). These could include ePortfolios or retrieval practice applications (e.g. Kahoot!). This desire to diversify assessment approaches at post-primary level is particularly relevant within the Irish context as demonstrated by recent discussions at government level (O'Brien 2021). While the relevance of this research would have been enhanced by examining test-takers' perceptions of a broader range of technology-based assessments, CBEs are likely to still play an important role in the future, particularly if they can alleviate criticisms common to their paper-based equivalents. Although qualitative research does not aim to be generalisable (Salomon 1991), the small sample size of the current research should also be considered when interpreting this study's results.

These findings also offer some clear directions for future research. For example, further research regarding the readiness of post-primary test-takers for digital assessments is needed, particularly regarding test-takers' digital literacy. Fraillon, Schulz, and Ainley (2013) define digital literacy as the ability to use digital resources to collect, create, transform, and safely use information. While post-primary students are often erroneously considered to be 'digital natives' (Prensky 2001), research has found that despite early and prolonged exposure to technology, they often lack the skills necessary for effective and critical technology use (e.g. Lazonder et al. 2020). This aligns with the data gathered here whereby participants acknowledged their difficulties in generalising behaviours across digital environments and self-identified limitations in their digital literacy skills (e.g. typing proficiency). The findings from this study reveal a new and fertile ground for future research particularly regarding the use of designated training programmes for digital literacy skills and whether they could support readiness for and performance in CBEs.

From a methodological perspective, the current study highlights the value of including *test-taker* voice in assessment research, reflecting progress in learning research with learner voice (e.g. Flynn 2017). The test-takers in this research highlighted what they required of a CBE's interface, including 'warnings' if a question had been forgotten, the freedom to navigate between items and the ability to review their final responses. Given the insights obtained from the participants involved in this research, the co-production of high stakes CBEs *with* post-primary students could be particularly beneficial to those countries, like Ireland, who are only beginning to explore the wider use of CBEs for post-primary students. Research that examined the co-production of health interventions in English post-primary schools (Ponsford et al. 2021), demonstrated that post-primary students are well placed to highlight facilitators or barriers to implementation and acceptability. They can also identify potential unintended consequences and ways of addressing these. Adopting this approach could be very valuable to those involved in the design and development of CBEs.

Conclusion

Given the context in which this study took place, and taking into consideration the recent initiatives involving TBAs for the Leaving Certificate Examination (SEC, 2021) as well as the upcoming revisions to the 'Digital Strategy for Schools' (DES, 2021), the findings of this research will be particularly pertinent to Irish educational policy makers. However, they also have relevance well beyond the Irish context. It is hoped that these findings can be used as a tool for stakeholders to reflect on the design of CBEs so that they can be used to maximum effect.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Note

1. That some participants completed items with static multimedia objects, whilst others completed items with dynamic objects reflects the fact that the larger piece of research from which this study is taken from was focused on the comparison of test-taker performance and attentional behaviour across these two different types of multimedia stimuli.

Notes on contributors

Paula Lehane is an Assistant Professor in the School of Inclusive and Special Education in the Institute of Education at Dublin City University (DCU). She previously worked as a primary school teacher and was the Special Educational Needs (SEN) coordinator of a large urban primary school. She recently completed her doctoral research on the design of digital assessments for post-primary students. Her research interests encompass assessment, inclusion and technology as they relate to the education systems.

Dr *Darina Scully* is an Assistant Professor of Child & Adolescent Learning and Development at DCU's Institute of Education. She holds a PhD in Psychology from Trinity College, Dublin, and she is currently lecturing in quantitative research methods and social, personal & health education. Her research interests span various assessment, teaching and learning issues in primary, post-primary and higher education contexts.

Professor *Michael O'Leary* holds the Prometric Chair in Assessment at DCU and is Director of the Centre for Assessment Research Policy and Practice (CARPE) at the Institute of Education there. He leads a programme of research at CARPE focused on assessment in education and in the workplace.

ORCID

Paula Lehane http://orcid.org/0000-0003-0856-3505
Darina Scully http://orcid.org/0000-0002-6076-717X
Michael O'Leary http://orcid.org/0000-0002-6771-904X

References

Alemdag, E., and K. Cagiltay. 2018. "A Systematic Review of Eye Tracking Research on Multimedia Learning." *Computers & Education* 125: 413–428.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Lanham, MD: American Educational Research Association.

Baker, R., and J. Clarke-Midura. 2013. "Predicting Successful Inquiry Learning in a Virtual Performance Assessment for Science." Paper presented at the 21st International Conference on User Modeling, Adaptation, and Personalization, Rome, June 10–14. doi:10.1007/978-3-642-38844-6_17.

Berry, R. 2011. "Assessment Reforms Around the World." In *Assessment Reform in Education: Policy and Practice*, edited by R. Berry and B. Adamson, 89–102. Dordrecht: Springer.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. doi:10.1191/1478088706qp063oa.

Bryant, W. 2017. "Developing a Strategy for Using Technology-Enhanced Items in Large-Scale Standardized Tests." *Practical Assessment, Research & Evaluation* 22 (1): 1–5. https://pareonline.net/getvn.asp?v=22&n=1.

Cantillon, P., B. Irish, and D. Sales. 2004. "Using Computers for Assessment in Medicine." *British Medical Journal (Clinical Research Edition)* 329 (7466): 606–609.

Chatterji, M. 2013. "Insights, Emerging Taxonomies, and Theories of Action Toward Improving Validity." In *Validity and Test Use: An International Dialogue on Educational Assessment, Accountability and Equity*, edited by M. Chatterji, 273–330. Bingley: Emerald Group Publishing.

Cheung, A., and R. Slavin. 2013. "The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis." *Educational Research Review* 9: 88–113.

Department of Education and Skills (DES). 2015. *Digital Strategy for Schools 2015-2020: Enhancing Teaching, Learning and Assessment*. https://www.education.ie/en/Publications/Policy-Reports/Digital-Strategy-for-Schools-2015-2020.pdf.

Department of Education and Skills (DES). 2017. *Digital Learning Framework for Post-Primary Schools*. https://www.education.ie/en/Schools-Colleges/Information/Information-Communications-Technology-ICT-in-Schools/digital-learning-framework-secondary.pdf.

Department of Education and Skills (DES). 2020. *Digital Learning 2020: Reporting on Practice in Early Learning and Care, Primary and Secondary Contexts*. https://assets.gov.ie/78007/a37413f9-7423-44bf-86df-01762e04a408.pdf.

Department of Education and Skills (DES). 2021. *Digital Strategy for School Consultation Framework* [website]. https://www.education.ie/en/Schools-Colleges/Information/Information-Communications-Technology-ICT-in-Schools/digital-strategy-for-schools-consultation-framework.html.

Deutsch, T., K. Herrmann, T. Frese, and H. Sandholzer. 2012. "Implementing Computer-Based Assessment—A web-Based Mock Examination Changes Attitudes." *Computers & Education* 58 (4), 1068–1075. https://www.learntechlib.org/p/67493/.

Donnolly, K. 2021. "Leaving Cert Computer Science a 'Fair Exam with Some Challenges'". *Irish Times*, May 22. https://www.independent.ie/irish-news/education/leaving-cert-computer-sciencea-fair-exam-with-some-challenges-40456899.html.

Elbabour, F., O. Alhadreti, and P. Mayhew. 2017. "Eye Tracking in Retrospective Think-Aloud Usability Testing: Is There Added Value?" *Journal of Usability Studies* 12 (3): 95–110. http://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Elbabour_May2017.pdf.

Flynn, P. 2017. *Embedding Student Voices in Education Discourse: Curricular co-Construction and Development*. Dublin: National Council for Curriculum and Assessment (NCCA). https://ncca.ie/media/3442/16539-ncca-the-learner-voice-research-study-v2.pdf.

Fraillon, J., W. Schulz, and J. Ainley. 2013. "International Computer and Information Literacy Study: Assessment Framework." International Association for the Evaluation of Educational Achievement (IEA).

Garrett, H. 1937. *Statistics in Psychology and Education*. New York: Longmans.

Harms, M., and J. Adams. 2008. "Usability and Design Considerations for Computer-Based Learning and Assessment". Paper Presented at American Educational Research Associations (AERA), New York. https://pdfs.semanticscholar.org/3661/cbe8b6d8fec7ad37648164d02f2a80d47960.pdf.

Hochlehnert, A., K. Brass, A. Moeltner, and J. Juenger. 2011. "Does Medical Students' Preference of Test Format (Computer-Based vs. Paper-Based) Have an Influence on Performance." *BMC Medical Education* 11: 89. doi:10.1186/2F1472-6920-11-89.

Hoffler, T., and D. Leutner. 2007. "Instructional Animation vs. Static Pictures: A Meta-Analysis." *Learning and Instruction* 17 (6): 722–738.

Huff, K. L., and S. G. Sireci. 2001. "Validity Issues in Computer-Based Testing." *Educational Measurement: Issues and Practice* 20 (3): 16–25. https://www.learntechlib.org/p/94639/.

Ilomäki, L., and M. Lakkala. 2018. "Digital Technology and Practices for School Improvement: Innovative Digital School Model." *Research and Practice in Technology Enhanced Learning* 13: 1–32.

Jordan, K. 1998. "Defining Multimedia." *IEEE Multimedia* 5: 8–15.

Karakolidis, A., M. O'Leary, and D. Scully. 2021. "Animated Videos in Assessment: Comparing Validity Evidence from and Test-Takers' Reactions to an Animated and a Text-Based Situational Judgement Test." *International Journal of Testing* 21 (2): 57–79. doi:1080/15305058.2021.1916505.

Lazonder, A. W., A. Walraven, H. Gijlers, and N. Janssen. 2020. "Longitudinal Assessment of Digital Literacy in Children: Findings from a Large Dutch Single-School Study." *Computers & Education* 143: 103681.

Lee, Y. H., J. Hao, K. Man, and L. Ou. 2019. "How Do Test Takers Interact with Simulation-Based Tasks? A Response-Time Perspective." *Frontiers in Psychology* 10: 906.

Lehane, P. 2019. *Leaving Certificate Computer Science: Factors to Consider When Developing Computer-Based Examinations*. Dublin: National Council of Curriculum and Assessment. https://www.ncca.ie/media/4081/lccs_cbe_factorstoconsider_lehane2019-for-ncca-website.pdf.

Levy, R. 2012. "Psychometric Advances, Opportunities, and Challenges for Simulation-Based Assessment." Paper Presented at the Invitational Research Symposium on Science Assessment, May 2012. https://www.ets.org/Media/Research/pdf/session2-levy-paper-tea2012.pdf.

Lindner, M., A. Eitel, J. Barenethien, and O. Köller. 2021. "An Integrative Study on Learning and Testing with Multimedia: Effects on Students' Performance and Metacognition." *Learning and Instruction* 71: 101100.

Lindner, M., O. Lüdtke, S. Grund, and O. Köller. 2017. "The Merits of Representational Pictures in Educational Assessment: Evidence for Cognitive and Motivational Effects in a Time-on-Task Analysis." *Contemporary Educational Psychology* 51: 482–492.

Mayer, R. 2008. "Applying the Science of Learning: Evidence Based Principles for the Design of Multimedia Instruction." *American Psychologist* 63 (8): 760–769. doi:10.1037/0003-066X.63.8.760.

Mayer, R. 2014. "Cognitive Theory of Multimedia Learning." In *The Cambridge Handbook of Multimedia Learning*. 2nd ed., edited by R. E. Mayer, 43–71. Cambridge: Cambridge University Press.

Mayer, R., and R. Moreno. 2002. "Animation as an Aid to Multimedia Learning." *Educational Psychology Review* 14 (1): 87–99.

Messick, S. 1994. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Researcher* 23 (2): 13–23.

Mitra, D. 2018. "Student Voice in Secondary Schools: The Possibility for Deeper Change." *Journal of Educational Administration* 56 (5): 473–487. doi:10.1108/JEA-01-2018-0007.

New Zealand Qualification Authority (NZQA). 2021. *2021 Examination Timetable* [webpage]. https://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/exams-and-portfolios/examination-timetable.pdf.

New Zealand Qualifications Authority (NZQA). 2018. *Digital Assessment: NCEA Digital Trials and Pilots* [webpage]. https://www.nzqa.govt.nz/about-us/future-state/digital-assessment-trials-pilots/.

O'Brien, C. 2021. "Leaving Cert has 'pernicious' Impact on Learning, Committee Hears." *Irish Times*, November 16. https://www.irishtimes.com/news/education/leaving-cert-has-pernicious-impact-on-learning-committee-hears-1.4730570.

Olsen, A., L. Smolentzov, and T. Strandvall. 2010. "Comparing Different Eye Tracking Cues When Using the Retrospective Think Aloud Method in Usability Testing." Paper presented at the Journal of Usability Studies, British Computer Society, September, 45–53. https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2010.8.

Organisation for Economic Co-operation and Development (OECD). 2015a. *Students, Computers and Learning: Making the Connection.* https://doi.org/10.1787/9789264239555-en.

Organisation for Economic Co-Operation and Development (OECD). 2015b. *PISA 2015 Released Field Trial Cognitive Items.* https://www.oecd.org/pisa/test/PISA2015-Released-FT-Cognitive-Items.pdf.

Organisation for Economic Co-Operation and Development (OECD). 2017. *PISA: 2015 Technical Report.* https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf.

Parshall, C., and J. Harmes. 2008. "The Design of Innovative Item Types: Targeting Constructs, Selecting Innovations, and Refining Prototypes." *CLEAR Exam Review* 19 (2): 18–25. https://www.clearhq.org/resources/CLEAR_summer08_4.pdf.

Ponsford, R., R. Meiksin, S. Bragg, J. Crichton, L. Emmerson, T. Tancred, N. Tilouche, et al. 2021. "Co-production of Two Whole-School Sexual Health Interventions for English Secondary Schools: Positive Choices and Project Respect." *Pilot & Feasibility Studies* 7: 50. https://pilotfeasibilitystudies.biomedcentral.com/articles/10.1186/s40814-020-00752-5.

Prensky, M. 2001. "Digital Natives, Digital Immigrants: Part 1." *On the Horizon* 9 (5): 1–6.

QSR International. 2020. "NVivo 12." http://www.qsrinternational.com/products_nvivo.aspx.

Russell, M. 2016. "A Framework for Examining the Utility of Technology-Enhanced Items." *Journal of Applied Testing Technology* 17 (1): 20–32. http://www.jattjournal.com/index.php/atp/article/view/89189/67798%5Cnhttp://www.jattjournal.com/.

Salkind, N. J. 2010. *Encyclopedia of Research Design.* Thousand Oaks, CA: SAGE. doi:10.4135/9781412961288.n460.

Salomon, G. 1991. "Transcending the Qualitative-Quantitative Debate: The Analytic and Systemic Approaches to Educational Research." *Educational Researcher* 20 (6): 10–18.

Scully, D. 2017. "Constructing Multiple-Choice Items to Measure Higher-Order Thinking." *Practical Assessment, Research, and Evaluation* 22: 4.

Shiel, G., C. Kelleher, C. McKeown, and S. Denner. 2016. "Future Ready? The Performance of 15-Year-Olds in Ireland on Science, Reading Literacy and Mathematics in PISA 2015." Educational Research Centre.

Siozos, P., G. Palaigeorgiou, G. Triantafyllakos, and T. Despotakis. 2009. "Computer Based Testing Using "Digital ink": Participatory Design of a Tablet PC Based Assessment Application for Secondary Education." *Computers & Education* 52 (4): 811–819.

Starkey, L., M. Shonfeld, S. Prestridge, and M. Gisbert Cervera. 2021. "Special Issue: Covid-19 and the Role of Technology and Pedagogy on School Education During a Pandemic." *Technology, Pedagogy & Education* 30 (1): 1–5.

State Examination Commission. 2021. *Leaving Certificate Computer Science* [webpage]. https://www.examinations.ie/?l=en&mc=ex&sc=cs.

Tamim, R. M., R. M. Bernard, E. Borokhovski, P. C. Abrami, and R. F. Schmid. 2011. "What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study." *Review of Educational Research* 81 (1): 4–28.

Teig, N., R. Scherer, and M. Kjærnsli. 2020. "Identifying Patterns of Students' Performance on Simulated Inquiry Tasks Using PISA 2015 log-File Data." *Journal of Research in Science Teaching* 57 (9): 1400–1429.

Tuzinski, K. 2013. "Simulations for Personnel Selection: An Introduction." In *Simulations for Personnel Selection*, edited by M. Fetzer and K. Tuzinski, 1–16. New York: Springer.

Walker, R., and Z. Handley. 2016. "Designing for Learner Engagement with Computer-Based Testing." *Research in Learning Technology* 24: 1–14.

Weick, K. E. 1995. *Sensemaking in Organizations*. Thousand Oaks, CA: Sage Publications.

Wise, S. 2019. "Controlling Construct-Irrelevant Factors Through Computer-Based Testing: Disengagement, Anxiety, & Cheating." *Education Inquiry* 10 (1): 21–33.