Tan, S., Wiebrands, M., O'Halloran, K. L. & Wignell, P. (2020). Analysing Student Engagement with 360-degree videos through Multimodal Data Analytics and User Annotations. *Technology, Pedagogy and Education,* 29(5), 593-612, DOI: 10.1080/1475939X.2020.1835708

Analysing student engagement with 360-degree videos through multimodal data analytics and user annotations

This paper presents the findings from a case study which explores how students engage with learning materials mediated through 360-degree panoramic videos. The case study extends findings from a project which developed an online platform for viewing and annotating 360-degree videos, and for providing data analytics (heatmaps, viewpoint tracking and area of interest displays) to map student engagement with video content. In this case, information obtained from the data analytics forms the basis for the analysis which, taking a social semiotic perspective, explores (a) what multimodal resources in the video-recorded classroom activity function to attract and hold students' attention, and (b) what multimodal resources students employ when interacting with the 360-degree videos and for which purpose, taking the context into account. The findings indicate that for students to engage meaningfully with educational content requires familiarity with the demands of 360 video technology, which for some users can prove distracting, or even disabling.

Keywords: 360-degree video technology; student engagement, data analytics; heatmaps, multimodality

Introduction

In response to rapid developments in video technology and applications over the past few decades, video-based approaches to teaching and learning have evolved to become a valuable tool in pre- and in-service teacher education worldwide. For example, video-based approaches have proven to be effective in fostering critical reflection, particularly in mathematics and science education (e.g. Chittleborough, Cripps Clark, & Chandler, 2015; Hackling, Murcia, Ibrahim-Didi, & Hill, 2014; van Es, Stockero, Sherin, Van

Zoest, & Dyer, 2015; van Es, Cashen, Barnhart, & Auger, 2017). Nonetheless, while technology-enhanced approaches to teaching and learning have the capacity to improve experiences for some students, research has shown that video-based teaching and learning methods can also present major challenges (e.g. Crook et al., 2012; Hung, 2016, 2018; Major & Watson, 2018; Sauli, Cattaneo, & van der Meij, 2018), particularly for off-campus, online learners. This includes issues such as disconnection and disengagement (e.g. Boling, Hough, Krinsky, Saleem & Stevens, 2012), as well as increased gradual reduction in attention given to instructional tasks when viewing videos (McGowan & Hanna, 2015).

In order to address some of these challenges, an interactive online platform which allows students to explore teaching and learning activities recorded with 360-degree video cameras was developed in the project "Encouraging and Mapping Student Engagement through 360-degree Video Annotation and Data Analytics", funded by the Australian Technology Network (ATN) (for detailed background information about the project, see O'Halloran, Tan, Wiebrands, Sheffield, Wignell, & Turner, 2018). The year-long project, undertaken at Curtin University, in collaboration with five other universities across Australia, aimed to (a) improve the quality of students' online experience and enhance critical engagement with course content by developing an interactive online platform for viewing and annotating 360-degree videos of learning activities, and (b) use data analytics in the form of information visualisations to map student engagement with video-recorded learning materials. While 360-degree video technology can be potentially applied in any discipline, the platform was initially trialled in mathematics and science teacher education programmes because the ability to both utilise and understand the affordances of multimodal resources and their contribution to meaning-making is of particular importance in these two subject areas, with implications for both students and teachers (e.g. see O'Halloran, 2011, 2015).

This paper presents the findings from a case study which explores how student participants enrolled in one such course at Curtin University engaged with two videos of learning activities mediated through 360-degree video technology. Data analytics in the form of heatmaps, viewpoint tracking and area of interest displays form the basis for the analysis which, taking a multimodal social semiotic perspective, explores what multimodal resources in the video-recorded classroom activities function to attract and hold students' attention, and what multimodal resources students employ when interacting with the 360-degree videos, and for which purposes. It considers the challenges students may have encountered, and the implications for teaching and learning with 360-degree video technology.

In what follows, we first provide the justification for our case study and the research questions that motivate it.

Study justification and motivations

360-degree video presents an emergent technology. Despite recent advances in the use of virtual reality and 360-degree video technology with a view to improving digital teaching and learning experiences (e.g. Argyriou, Economou, & Bouki, 2017; Assilmia, Yun, Okawa, & Kunze, 2017; Elmezeny, Edenhofer, & Wimmer, 2018; Gregory, Lee, Dalgarno, & Tynan, 2016; Hales & Kalyvaki, 2017), at present, it is generally not well understood exactly how students use the technology to engage with course content in such situations and the challenges this may pose (Alzahrani, Gardner, Callaghan, & Alrashidi, 2015; Georgiou & Kyza, 2017). Similarly, while multimodal learning analytics is a trending method for modelling student behaviour (e.g. Andrade, Delandshere, & Danish, 2016), measuring complex learning tasks (e.g. Blikstein &

Worsely, 2018), and identifying and tracing the physical movement and interactions of students and teachers during body-based learning activities (e.g. Healion, Russel, Cukurova, & Spikol, 2017; Smith, King, & Gonzales, 2016), its application in the context of modelling student engagement with 360-degree video technology from a multimodal social semiotic perspective is as yet unexplored and unprecedented.

To fill this gap, the case study presented in this paper explores how online students who participated in the above-mentioned project engaged with two videos of different learning activities recorded from a 360-degree perspective. According to Shernoff (2013) student engagement can be described as “heightened, simultaneous experience of concentration, interest, and enjoyment in the task at hand” (pp. 12). Engagement has also been linked to paying attention, which is a measurable behaviour that falls within the behavioural spectrum of student engagement (e.g. Gill & Remedios, 2013). For the purpose of this study, student engagement is viewed from a multimodal perspective and is concerned with the embodied ways in which student users interacted with the 360-degree videos, the spaces they viewed, and the actions they performed when navigating the 360-degree video environment. Information obtained from built-in data analytics (e.g. heatmaps, viewpoint tracking and area of interest displays), which provide empirical evidence about students’ interactions with the 360-degree videos, forms the basis for the subsequent analysis, which, conducted from a multimodal social semiotics perspective, seeks to address the following concomitant research questions:

- How is students’ engagement affected by different contextual factors in the 360-degree videos? For example, how do the multimodal resources deployed in the video-recorded classroom activity function to attract and hold students’ attention (e.g. teacher/students speaking, moving, looking, gesturing, handling material objects, use of classroom space); and how do these factors contribute to student engagement with the videos and the recorded classroom activities?

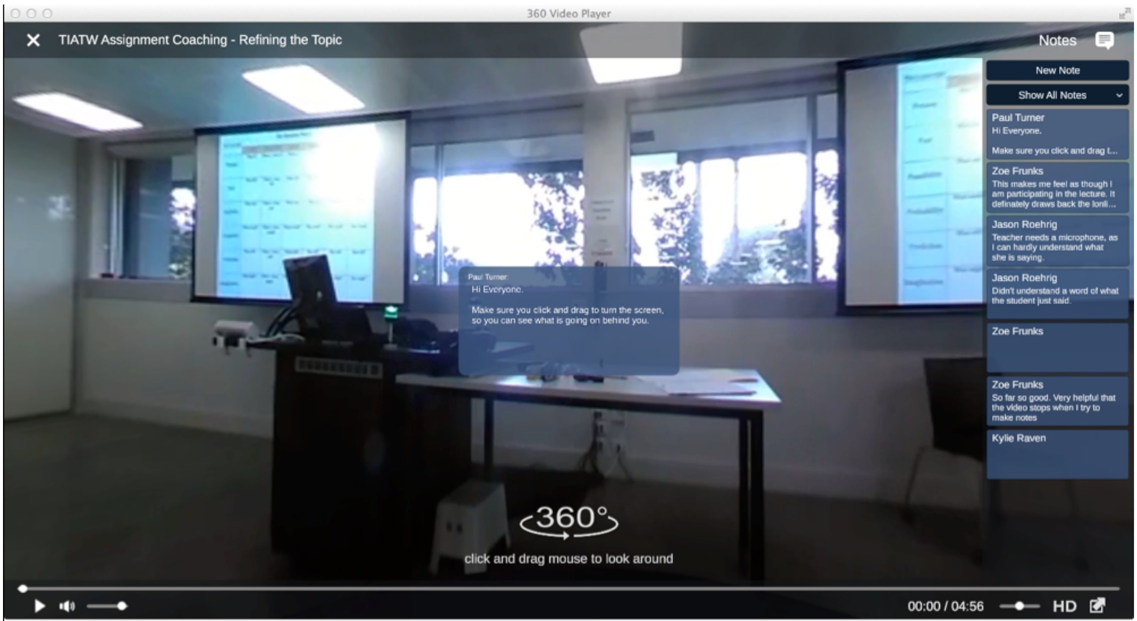
- What, multimodal resources do students employ when interacting with the 360-degree videos and for which purposes (e.g. navigating the 360-degree environment by rotating the screen/viewing angles, zooming in/out to focus on a particular scene or activity, using written language in annotations for communication or self-reflection), and how do the challenges and affordances of the technology influence the ways in which students interact with the videos and the content?

In what follows, we first provide an overview of the 360-degree video player and the in-built data analytics tools used for mapping student engagement. This is followed by a brief summary of the project scope and participants. We then describe the sample data used for the case study, and provide an outline of the multimodal social semiotic approach which has been adopted for interpreting the empirical results, obtained from the data analytics.

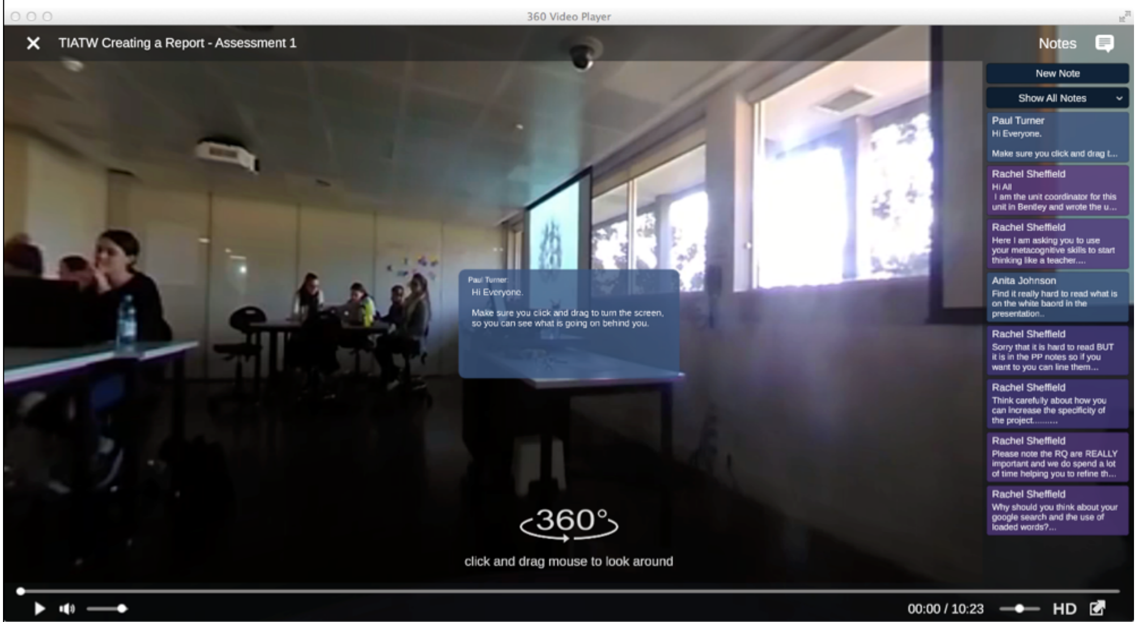
Materials and methods

Overview of 360-degree video player and data analytics tools

The 360-degree video platform, which was developed with the aims of improving the quality of students' online viewing experience and enhancing critical engagement with course content, consists of a video player which provides teachers and students with a panoramic view of classroom activities recorded with 360-degree video cameras. The video player includes a tool for annotating the video content by means of free-text notes that can be temporally and spatially positioned as overlays in the video. Notes can be used by lecturers to guide students or to set learning tasks and ask questions for students to engage with. Notes can also be used by students to respond to the set learning tasks, or for communicating with their teacher. Figure 1 presents a screenshot of the notes displayed at video start position (and the sidebar) for the two videos analysed in this paper.



(a) Annotation note displayed at video start position (Video 1)

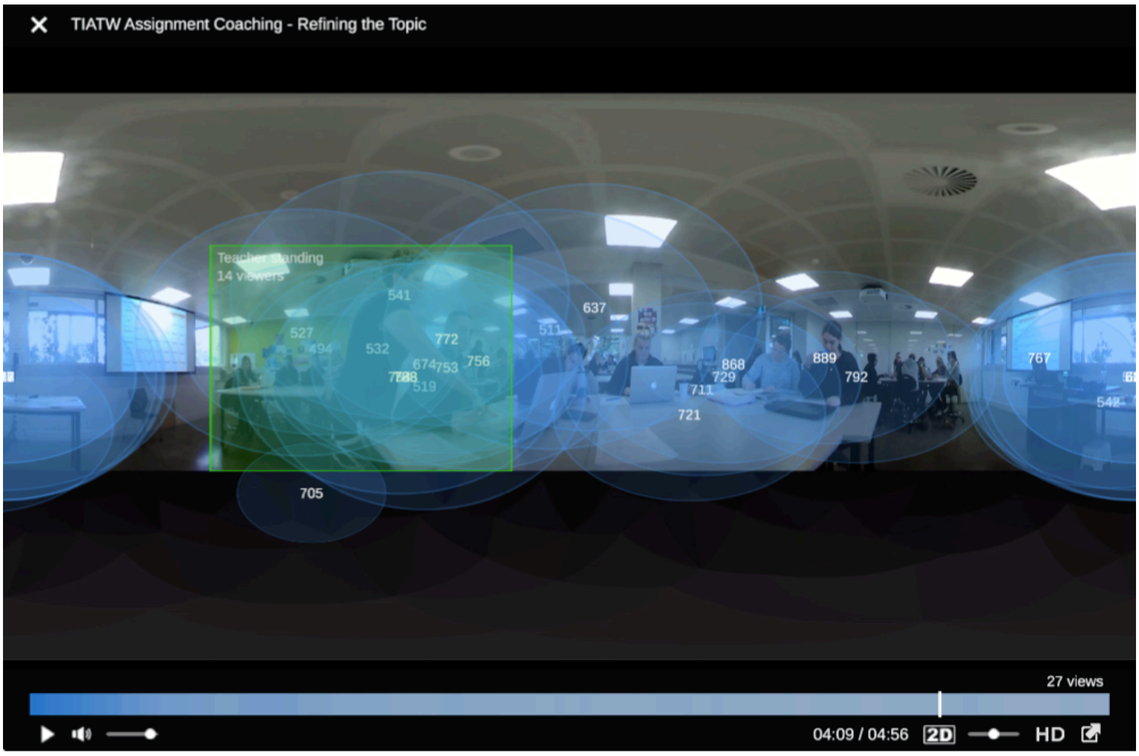


(b) Annotation note displayed at video start position (Video 2)

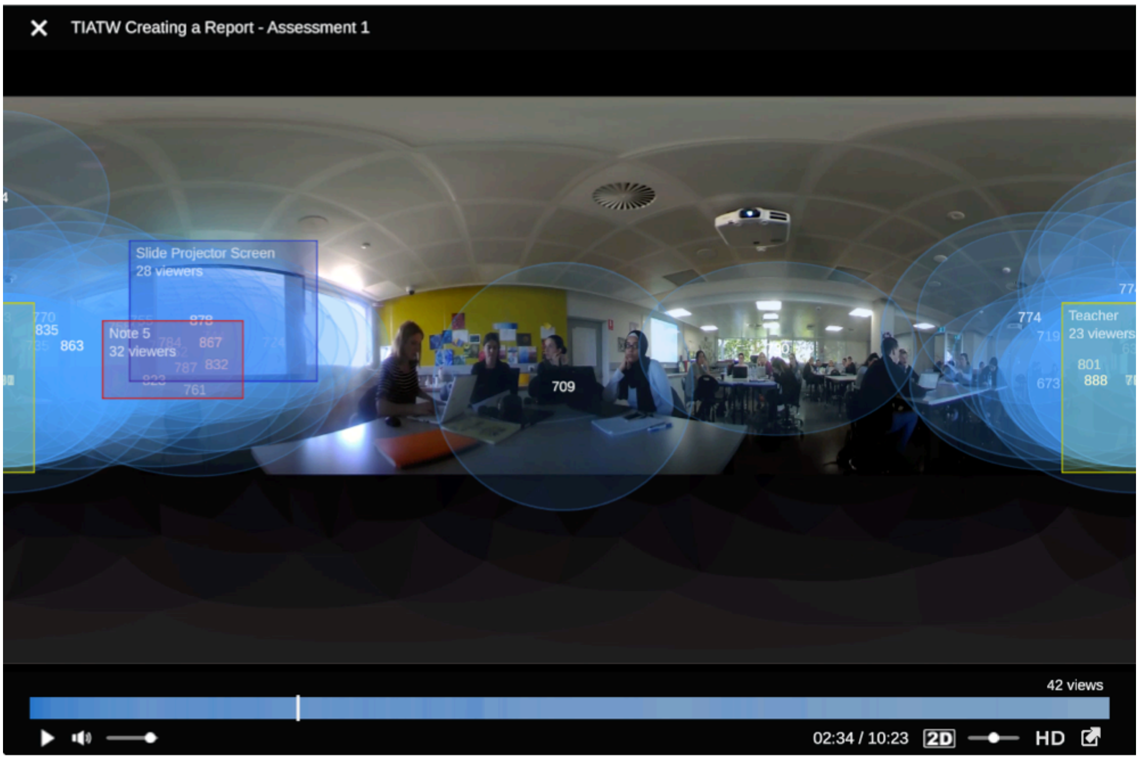
Figure 1. Annotation notes displayed at video start position – (a) Video 1, (b) Video 2.

Additional features for academics' use include a number of analytic tools for evaluating student engagement with video-recorded classroom events by means of information visualizations. For example, dynamic view displays provide information about where viewers were looking in the video and whether they resized their field of view by manipulating the player window or by zooming in or out to focus on a particular scene or activity. In addition, an 'area of interest' tool allows academics to create bounded areas of interest (displayed as rectangles) around certain activities or scenes that students are expected to watch (e.g. teacher talking, walking, writing on whiteboard), and which indicate how many students viewed these areas of interest while they were watching the video. In addition, the video seek bar at the bottom of the video player shows the number of active views at a particular playback time, and also provides an indication whether students actually watched the video from beginning to end.

Figure 2 shows examples of view displays with overlaid areas of interest which give an overview of the areas viewed by all student users at a particular point of time in the two analysed videos, whereby their respective field of view is indicated by the size of the blue circles or ellipses, while the numbers in the circles represent the individual student view IDs. Note that in the analytics video player the video is displayed in a flattened-out equirectangular format, that is, the left and right hand edges of the video viewer indicate the same space and starting point in the 360-degree spherical view as displayed in Figure 1.



(a) View display with area of interest – all student views (Video 1)



(b) View display with areas of interest – all student views(Video 2)

Figure 2. View display with overlaid areas of interest – all student views – (a) Video 1, (b) Video 2.

Lastly, different types of static and dynamic heatmaps overlaid on the analytics video viewer provide additional information about the spaces users explored when they interacted with video. Examples of heatmap visualisations that formed the basis for the analysis of viewing patterns are shown in Table 1 under 'Results'. In this case, aggregate (summed) heatmaps (left column in Table 1) display the locations individual students were viewing most throughout the duration of the video. Aggregate (maxima) heatmaps (right column in Table 1) display all the locations viewed by students throughout the entire video. The more intensely (red) coloured areas in the heatmaps indicate the areas where viewers were looking at most frequently in the video. (Note: The data analytics tools include two additional (dynamic) heatmaps not shown here: (i) temporal (local heat) for displaying the views recorded at a particular point of time as the video unfolds, and (ii) temporal (global heat) for displaying the distribution of views at a particular point in time in relation to the total number of views over the entire video.)

Data analytics can be displayed for all views, academic views, student views, single user views and single viewing sessions.

Project scope and participants

The 360-degree software application was trialled in mathematics and science teacher education programs taught at Curtin University (Perth, Western Australia) and five other universities across Australia. The participants in this study were pre-service teachers enrolled in these programs. The total number of students registered to the units from these programs was 644. Student participation in the project was voluntary. Students who registered to the units in which the 360-degree video platform was trialled received recruitment materials that informed them about the project. Students who accepted the invitation to participate in the project signed a consent form when they logged onto the online 360-degree video platform. Specific guidance or training was not provided. A user guide on how to use the 360-degree video player was made available on the project website (http://online360video.education/). For the duration of the project, 33 distinctive video files of classroom activities recorded with 360-degree cameras had been uploaded for ten units taught at Curtin University and other participating universities. Teachers experimented with lesson recordings and camera placement. Placing the camera where a student would be standing or sitting in the classroom was deemed to provide the best simulation of an immersive experience for viewers (O'Halloran, Tan, Wiebrands, Sheffield, Wignell, & Turner, 2018).

Data sample

The data for this case study comes from the OUA (Open Universities Australia) online science unit 'Teachers Inquiring About The World (SP4 2017)' for a teacher education course taught at Curtin University for which 476 students had registered. A total of ten videos were uploaded to this course, which received a total of 239 individual student video views. The case study presented in this paper investigates how individual student users interacted and engaged with two different videos from this course that attracted the highest number of student views.

The first video, entitled 'TIATW Assignment Coaching - Refining the Topic', (Video 1) is 04:56 minutes long. It was released on 26 October 2017, and first viewed on 20 November 2017. The total number of users (including academics) for Video 1 was 62, with a total of 79 viewing sessions, and an average number of viewing sessions per user of 1.3.

The second video from the same unit, entitled 'TIATW Creating a Report - Assessment 1' (Video 2), is 10:23 minutes long. It was released on 27 October 2017, and first viewed on 06 December 2017. The total number of users for Video 2 was 43, with a total of 76 viewing sessions, and an average number of viewing sessions per user of 1.8.

Video 1 contained a total of seven annotation notes: six appended by students (including two notes without text input) and one note appended by a teachers' aide (academic) at video start position to guide students on how to interact with the 360-degree video (see Figure 1). Video 2 contained a total of eight annotations: the same note at video start position, plus six notes appended by the course coordinator (i.e. the teacher featured in the two videos), and one student note. An extract of the written text in the annotation notes is provided in Table 2(a)–(b) in the section entitled 'Analysis and discussion of findings' below.

Analytical methods and interpretive approach

Data analytics for mapping student engagement with 360-degree videos

As a first step in the analysis, data analytics in the form of static and dynamic heatmaps (aggregate, summed and temporal) (for examples, see Table 1(a)–(b)), and dynamic view displays and areas of interests (as illustrated in Figure 2) were examined, compared and analysed for 59 individual student User IDs for Video 1 and 40 individual student User IDs for Video 2. Additional data was extracted from the 360-degree software application database in the form of tabulated data files, which provided empirical evidence about the action steps performed by individual students when they interacted with the video player (e.g. play, pause, stop, seek), whether and how often they had changed their view direction (captured in terms of rotational moves along the X-Y axis in the 360-degree spherical video environment) and whether they had adjusted their field of view by resizing their browser window or by zooming in or out on an activity.

Multimodal social semiotics for interpreting the results

The results obtained from the data analytics were then interpreted from a multimodal social semiotic perspective. Although initially and most fully developed for language, Halliday's (1978) social semiotic theory, which views language as but one semiotic resource among the many (e.g. spoken and written language, gaze, gesture, body posture, proxemics), has been influential in studies which approach teaching and learning as "*a dynamic process of sign making*" (Jewitt, Kress, Ogborn, & Tsatsarelis, 2001, p. 6, original italics.)

One of the key tenets of social semiotic theory is the 'metafunctional' principle which posits that language and other semiotic systems are structured in such a way to make three kinds of meanings simultaneously: (a) *ideational* meaning for construing our experience and knowledge of the world (i.e. experiential meaning) and for making logical connections in that world (i.e. logical meaning); (b) *interpersonal* meaning for enacting social relations and expressing attitudes; and (c) *textual* or *compositional* meaning for organising meanings into coherent messages relevant to their context (e.g. Halliday, 1978; Halliday & Matthiessen, 2014).

Social semiotic theory is essentially 'a theory of meaning as choice, by which language, or any other any other semiotic system, is interpreted as networks of interlocking options', whereby the particular choices that are made are not to be viewed as the result of conscious decisions but rather as unconscious choices from 'a set of possible alternatives' (Halliday, 1994, pp. xiv–xxvi). A social semiotic approach builds upon the assumption that meaning arises as a consequence of the choices made in any context, and that meaning is multiple (e.g. Jewitt et al., 2001).

Moreover, from a multimodal social semiotic perspective, different semiotic resources are perceived to have different affordances and constraints with regards to what can and what cannot not easily be expressed or represented in a given mode (e.g. Jewitt, 2003, 2012; Morell, 2015, 2017; Ryan, Scott, & Walsh, 2010; Sakr, Jewitt, & Price, 2016; Twiner, Coffin, Littleton, & Whitelock, 2010).

As the analysis will show, the choices teachers make during the video-recorded lessons impact the ways in which student interact with the videos. Similarly, the choices students make when interacting with and navigating the 360-degree environment are also important, because if students are not looking at the spaces they should be viewing, then they may miss significant content presented in the lesson.

Another key concept adapted from Halliday's social semiotic theory is the notion of context as modelled through the three key dimensions of *field*, *tenor* and *mode*, also known as register theory (e.g. Eggins, 2005; Martin, 2002; Martin & Rose, 2007). Register theory describes the impact of the way semiotic resources are used in context in relation to the above-described metafunctions. From this perspective, *field* is concerned with the nature of the domestic or institutionalised activity that is going on and relates to the ideational metafunction. *Tenor* is concerned with the way social relations are enacted through power and solidarity and relates to the interpersonal metafunction. *Mode* is concerned with the role of language in discourse, that is, whether it is written or spoken, and the 'information flow' across different media or channels of communication (e.g. speech, writing, images, video) (e.g. Martin, 2002, p. 56; Martin & White, 2005, pp. 27–28) and relates to the textual or compositional metafunction. In register theory, genres are seen to be realised (in discourse) through the registerial configurations of tenor, field, and mode.

From a social semiotic perspective, genres – including multimodal ones – are characterised as configurations of meaning that are 'relatively stable and recurrent, and which unfold in stages that are aimed at achieving a particular goal' (Zappavigna &

Zhao, 2017, p. 240). According to Martin and White (2005, p. 32-33), genres are characterised as 'social because we participate in genres with other people; goal oriented because we use genres to get things done and feel a sense of frustration when we don't resolve our telos; staged because it usually takes us a few steps to reach our goals.' However, even established genres may transform and evolve, 'often in tandem with expansions in meaning potential that are afforded by new semiotic technologies' (Zappavigna & Zhao, 2017, p. 240), resulting in new and additional layers of genre configurations over and above those associated with traditional ones.

The notion of multiplicity of meaning has also been influential in interactional studies, especially those that focus on the interplay between human interaction, space and learning, and digitally mediated forms of communication (e.g. Ciekanski & Chanier, 2008; Leijon, 2016). As Ciekanski and Chanier (2008, pp. 168–169) observe, the context in such situations is multifaceted and polyfocal, due to the opportunities which such environments provide to participants for moving from one field to another. Citing Goodwin and Duranti (1992, p. 3), they note that 'the notion of context involves a fundamental juxtaposition of two entities: (1) the focal event; (2) a field of action within which that event is embedded'.

From a social semiotic perspective, the activity of engaging with 360 video technology for engaging with educational content is indeed multifaceted and polyfocal, which requires a certain level of competency in negotiating and consolidating complex layers of different multimodal genres. While 360 video technology can offer new opportunities to students beyond those experienced with traditional classroom genres, it can also present new challenges, as will be shown in the following sections.

Analysis and discussion of findings

In this section, we present the findings from the analysis. In order to arrive at a better understanding of the multiple factors that contributed to the ways in which students interacted with the 360-degree videos and the recorded learning activities, we first provide our analysis of the contextual factors pertaining to the use of semiotic resources evident in the two video recordings, and the resultant impact on the ways in which students interacted with the 360-degree videos and the recorded content. We then describe the patterns of student engagement derived from the data analytics, which is followed by a comment on the use of written annotations by the teacher and the students.

Impact of contextual factors on student engagement with 360-degree videos

Many existing studies that investigate the use of semiotic resources in 'real life' or computer mediated teaching and learning situations argue that multimodal competence (including the use of space and camera placement for video-recorded sessions) plays an important role in the creation of meaning in a classroom context (e.g. Blikstad-Balas, 2017; Davidsen & Vanderlinde, 2014; Jewitt, 2012; Leijon, 2016; Lim, O'Halloran, & Podlasov, 2012; Morell, 2017; Tan, O'Halloran, & Wignell, 2016). Leijon (2016), for example, views space as a multimodal resource that plays an important role in the meaning-making process and forms an integral part of the setting in the creation of a learning sequence (Leijon 2016, p. 93-94). Lim, O'Halloran and Podlasov (2012, p. 235) similarly propose that the utilisation of classroom space, in combination with other semiotic resources (e.g. language, gesture and use of teaching materials), formalises the respective registers and microgenres found in a classroom context. They establish four different patterns in the use of classroom space for different purposes: (a) authoritative space, where the teacher is positioned in front of the teacher's desk and in the front centre of the classroom to conduct formal teaching and to provide instructions to facilitate the lesson; (b) interactional space, where the teacher is standing alongside the students' desks or between the rows of students' desks for personal consultation, or to offer guidance on the set task or clarification on an earlier instruction; (c) supervisory space, where the teacher moves alongside the students' desks, or up and down the side of the classroom, primarily for the purpose of supervision during student activities; and (d) personal space, here the teacher is sitting or standing behind the teacher's desk to pack and prepare for the next stage of the lesson.

As in Lim et al.'s (2012) study, the spatial layout of the classroom in which the video recordings of the two lessons were carried out is identical. The context (as represented, for example, by the teacher's use of space, recorded actions, camera placement and projected field of view as dislayed at video start position), however, is different in each situation. A diagrammatic representation of the context for the two video recordings is shown in Figure 3.

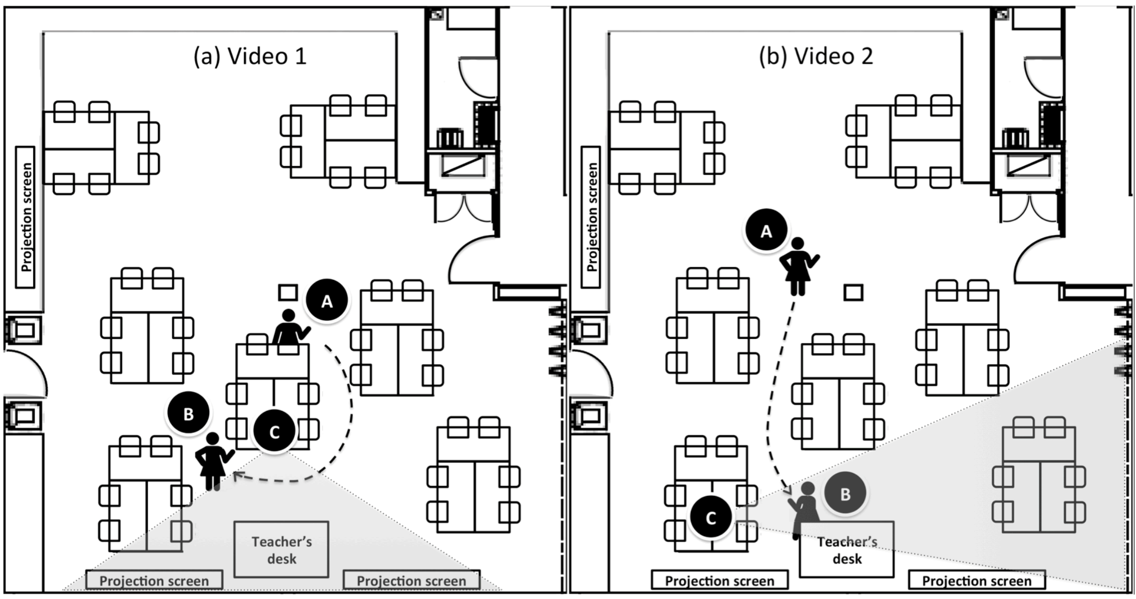


Figure 3. Classroom layout with teacher position (marked A, B), camera position (marked C) and projected field of view at video start position (area shaded in grey) – (a) Video 1, (b) Video 2.

To explain, the recorded classroom activity in Video 1 starts with the teacher sitting down with a group of students (marked A in Figure 3(a)) to discuss a chosen assignment topic with a female student (00:00 to 03:48 in the video), while the other students work away independently at their desks. At 03:49 to 03:52, the teacher gets up and moves across the room to attend to another student (see dotted line Figure 3(a)). She then stands next to a male student (marked B in Figure 3(a)) to discuss and give advice on a topic he had chosen for the assignment (03:53 to 04:56). In terms of Lim et al.'s (2012) interpretation, the teacher in Video 1 occupies interactional space, which – for the students' in the classroom – 'facilitates interaction' (Lim et al., 2012, p. 238). In terms of the other semiotic resources used in the recorded classroom activity sequence, it is mostly the teacher who speaks and who orchestrates the lesson sequence through her movements and actions. Although the students can be heard asking question and responding to the teacher, their voice is low and barely audible over the microphone. [Note: Incidentally, the lack of auditory feedback is a topic of concern communicated in a student's note; see rows 4 and 5 in Table 2(a)). It also highlights the importance of pertinent audio-visual cues to effectively guide users' attention in immersive video environments (e.g. Sarker, 2017; Salselas & Penha, 2019; Sheikh, Brown, Watson, & Evans, 2016).] Also, while there is written text displayed on the slide projection screen, it is static, and seemingly not pertinent to the recorded lesson sequence, as neither the teacher nor the students referred to it in the video recording or in the annotations.

For Video 1, the 360-degree camera was placed on the desk-space the teacher was occupying together with the students (marked C in Figure 3(a)). The projected field of view at video start position (that is, the initial scene online students would see when they started the video playback, depending upon the size and configuration of their browser window and computer screen) showed the teacher's desk facing two windows, flanked on either side by a slide projection screen (indicated by the area shaded in grey in Figure 3(a) and as illustrated in Figure 1(a)). That is, the projected field of view at video start position for Video 1 showed an area that remained unutilised throughout the entire video. It required that students interact with the 360-degree environment by rotating the screen to engage with the recorded classroom activity.

As outlined in Figure 3(b), the situational context for Video 2 is different. To begin with, the camera was placed on a different desk (marked C in Figure 3(b), with the projected field of view at video start position trained obliquely towards the teacher's desk (indicated by the area shaded in grey in Figure 3(b) and as illustrated in Figure 1(b)). In Video 2, the recorded activity starts with the teacher walking briskly from the centre of the room (that is, supervisory or interactional space, depending on the preceding classroom activity) towards the teacher's desk (00:00 to 00:03; see dotted line from position A to position B in Figure 3(b)). As all this happens very quickly within a few seconds at the beginning of the video, from the viewpoint of a student watching the video this means that the teacher walks directly into the student's field of view at video start position. The teacher stands briefly at the desk with her back towards the students (00:04 to 00:06), and then sits down on top of the teacher's desk, where she remains seated for the remainder of the video (00:07 to 10:23). Although her animated disposition, body posture and gestures suggest tempered informality, in terms of Lim et al.'s (2012) spatial classifications, the teacher occupies authoritative space, which 'constructs a formal tenor in the relationship between teacher and students' (Lim et al., 2012, p. 238), in this case not only for the students in the classroom but also for those watching the video recording. In terms of other semiotic resources (or modes) deployed in the orchestration of the video-recorded lesson in Video 2, it is again the teacher who speaks most of the time, although students can be heard asking questions and providing choral responses to her prompts and elicitations. The slide projection screen – towards which the teacher turns and gestures and points her clicker in regular intervals as the video unfolds – displays written text which appears to be tightly integrated in the lesson, and which aligns closely with her verbal utterances.

As the analysis will show, the different contextual factors and combinations of multimodal resources deployed in the two video recordings had a significant impact on the ways in which students engaged with the videos and the recorded classroom activities.

Mapping student engagement through data analytics

Patterns derived from the data analytics (i.e. heatmaps, view displays and areas of interest), together with the data of user behaviour extracted from the backend system, allowed us to draw inferences about the ways in which individual student users interacted with the two videos, and for which purposes. In summary, four main patterns of user engagement emerged: (1) students who were seemingly not adept at using 360 video technology; (2) students who engaged with the technology only to explore the 360-degree environment; (3) students who utilised 360 video technology to engage with the lesson content; and (4) students who utilised the technology to engage with the lesson content, but who also explored the 360-degree environment. Examples of typical heatmap patterns that allowed inferences to be drawn about user engagement are shown in Table 1(a)–(b). The results are expounded below.

Table 1(a). Examples of heatmap patterns for mapping student engagement – Video 1.

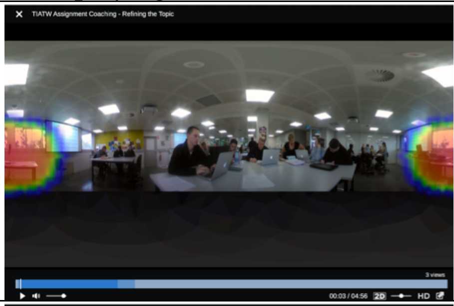
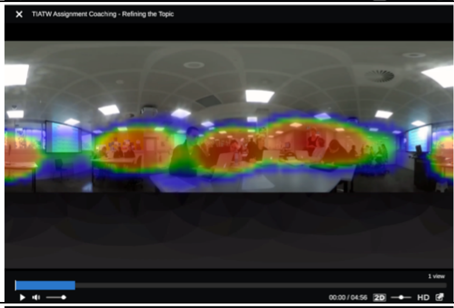
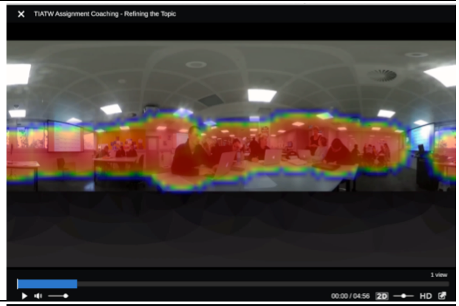
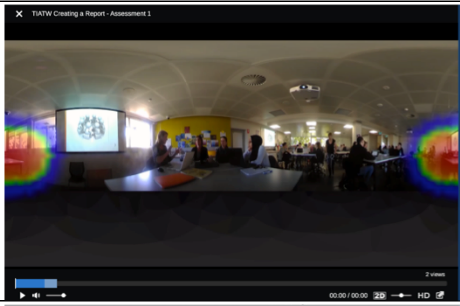
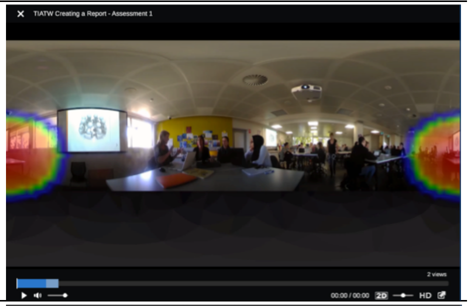
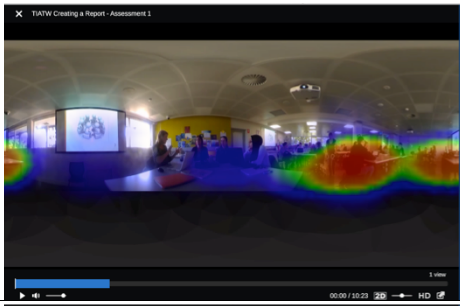
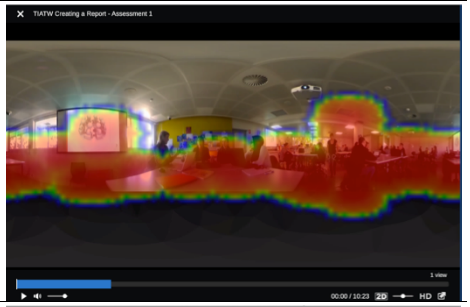
Types of user engagement	Heatmap patterns Aggregate (summed), displaying most viewed areas	 Aggregate (maxima), displaying total view coverage
1. Not adept at using 360 video technology		
2. Engaged with 360 video technology only		
3. Engaged with lesson content		
4. Engaged with lesson content, but also explored 360 space		

Table 1(b). Examples of heatmap patterns for mapping student engagement – Video 2.

Types of user engagement	Heatmap patterns Aggregate (summed), displaying most viewed areas	 Aggregate (maxima), displaying total view coverage

1. Not adept at using 360 video technology		
2. Engaged with 360 video technology only		
3. Engaged with lesson content		
4. Engaged with lesson content, but also explored 360 space		

Perceived patterns of student engagement with 360-degree video technology

1. Users who are not adept at using 360 video technology, or who might have experienced technical difficulties. A comparison of aggregate (summed) and aggregate (maxima) heatmaps (e.g. see heatmap patterns for row 1 in Table 1(a)–(b)) suggested that this group of students (14 users or 25% for Video 1; 5 users or 13% for Video 2) attempted to engage with the 360-degree video like they would with a conventional video: that is, they did not interact with the 360-degree video screen via mouse or tack

pad movements to change their view direction. Most of these users, even those who watched only briefly (9 users, watching on average only 4% of the video for Video 1; 5 users, watching on average 8% of the video for Video 2), nevertheless performed a series of action steps (such as play, pause, stop). Many also adjusted their field of view, e.g. by manipulating their viewer window or by zooming in and out. Although many users in this category quickly gave up after their first attempt, one student (for Video 1) viewed the video four times and performed a total of 42 recorded actions steps. In order to engage meaningfully with the educational content, the situational context for Video 1 required users to change their view direction. However, the data showed that five users watched the video from start to finish with the field of view as displayed at video start position (see Figure 1(a)). While they would have had auditory feedback, their view was directed at a classroom scene where nothing happened during the entire video. [Note: This group includes one user who performed only two actions steps (play and stop). Here the possibility exists that this user may have pressed the video play button but then turned away from the computer screen to attend to other matters, and thus not be engaged with the video or its content.] It needs to be clarified, however, that the majority of these unsuccessful interactions occurred before an annotation note was put in place that alerted students to the requirement to click and drag the screen to look around the 360-degree space (e.g. see Figure 1(a); for note text, see Table 2(a)–(b)).

2. Users who engaged with 360 video technology only. Here, the heatmaps and view displays indicated that these students (15 users or 26% for Video 1; 8 users or 21% for Video 2) explored most areas in the 360-degree video, often within a very short timeframe. The majority of these users did not view the video from beginning to end (on average watching only 16% of the video for Video 1; and 43% of the video for Video 2). However, they performed many directional moves and frequently deployed

other resources such as zooms. The data analytics further revealed that none of these users focused on the expected areas of interest (e.g. teacher in discussion with students for Video 1; teacher and slide projection screen for Video 2) long enough to be considered sufficiently engaged with the educational content of the recorded learning activity. In case of Video 2, the possibility exists that this type of behaviour may have been prompted by the written command (e.g. see row 1 in Table 2(a)–(b)) in the annotation note at video start position ‘Make sure you click and drag to turn the screen, so you can see what is going on behind you’ which – if taken literally – functions like an invitation to explore the 360-degree environment. Inadvertently, it may have encouraged some students to exploit the affordances of the technology only, without paying attention to the educational content.

3. Users who utilised 360 video technology to engage with the lesson content. The data analytics for this group of students (12 users for Video 1; 11 users for Video 2; accounting for 28% in both cases) showed that these users interacted with the 360-degree video screen explicitly to view the expected areas of interest. For Video 1, this group of users was, for most of the time, firmly focused on the teacher and her actions. Moreover, they also executed the required action steps via mouse movements or track pad manipulations to rotate the screen to track the teacher’s movements in interactional space. From the data available it can thus be inferred that these students were the ones who managed to consolidate the demands of the technology for the purpose of engaging with the educational content of the video-recorded materials, which in this case enabled them to experience the classroom activities as observes. The different contextual factors operating in Video 2 (as explained in the subsection on contextual factors above) had several implications for user engagement. For instance, the projected field of view at video start position for Video 2 (as illustrated in Figure 3(b)), combined with the

teacher's utilisation of classroom space, allowed students to engage with the educational content of the recorded lesson without having to navigate the 360-degree space, moreover as most of the annotation notes appended by the teacher, which reinforced the formal nature of teaching and learning activity, also appeared in that space. In Video 2 students had to stay focussed on the area designated as authoritative space (see Lim, O'Halloran & Podlasov, 2012) to engage with the educational content. The information provided by the data analytics showed that most users in this category nevertheless adjusted their view direction (if only minimally) – possibly motivated by the teacher's body posture and gestures – to shift their attention at intervals away from the teacher towards the written text displayed on the slide projection screen. It can be surmised that these students utilised the technology to access an area of interest which they would otherwise not have had access to had the lesson been recorded with a traditional video camera. Indeed, their interactions with the 360-degree video player afforded them to experience the recorded lesson environment in a similar manner as students in the classroom, oscillating between paying attention to the teacher and the written information displayed on the slide projection screen.

4. Users who utilised the technology to engage with the lesson content, but who also explored the 360-degree environment. The data analytics for this group showed that these students engaged with the video to view the expected areas of interest for a period of time. However, for a considerable number of students (12 users or 21% for Video 1; 15 users or 38% for Video 2), the affordances of the technology also proved a distraction, as in the course of the lesson, they moved away from the areas of interest to explore other areas of the 360-degree environment. In some instances, even parts of the floor and ceiling were explored (e.g. see heatmap patterns in row 4 in Table 1(a)). It is also possible that the exploration of space in these cases resulted from student boredom,

as only the teacher's voice can be heard clearly and no other activities were being carried out in the classroom simultaneously. The implication is the same: the exploration of the 360-degree space resulted in distraction, directing students' attention away from the main content of the lesson delivered verbally by the teacher.

Use of annotation notes

The analysis of the video data also revealed that the different contextual factors pertaining to the video recordings also impacted the ways and purposes for which students used the notes function. While it was originally envisioned that students would use annotations as a tool for reflecting critically on the video-recorded learning activities, or for responding to the learning tasks set by the teacher, an analysis of the text in students' notes showed that they did not use the annotations in the expected manner.

From a social semiotic perspective, language is seen as a tool for enacting social relations. Halliday (1994, p. 95) recognises several minor speech functions for realising interpersonal functions, such as greetings and exclamations, but the most fundamental types of speech role are seen to be concerned with giving and demanding actions or information, whereby exchanges involving actions are realised by offers and commands, and exchanges of information are realised by statements and questions (Halliday, 1994, pp. 68–69). Speech function is concerned with meaning as exchange, so each of these primary speech functions is complemented by set of expected responses or discretionary alternatives. That is, accepting or rejecting an offer, carrying out or refusing a command, acknowledging or contradicting a statement, answering or disclaiming a question (Hallliday, 1994, p. 69).

While the teacher (in Video 2; Video 1 contains no teacher notes) used a whole repertoire of speech functions in her notes to engage students linguistically with the field of teaching and learning, students apparently did not engage with the educational content in the teacher's notes (at least not overtly). Instead, in both videos, students used notes only as a means of conveying information pertaining to the field of technology, that is, either to provide feedback about their experience with the 360-degree video platform, or to report technical problems or issues. Also, while the teacher used the notes facility to communicate directly with a student by offering a solution to a reported problem (e.g. see rows 4 and 5 in Table 2(b)), students did not use annotations as a tool for dialogic interaction with the teacher (or other students).

Table 2(a)–(b) offers a comparison of students' and academics' use of annotation notes in the analysed videos.

Table 2(a). Comparison of students' and academics' use of annotation notes – Video 1.

User type	Text in annotation note	Field	Purpose	Predominant speech function
1. Academic	Hi Everyone. Make sure you click and drag to turn the screen, so you can see what is going on behind you.	Technology	Greeting Give advice	Greeting Demand action: command
2. Student A	This makes me feel as though I am participating in the lecture. It definitely draws back the loneliness curtain. Grateful for the opportunity to take part in this!	Technology	Provide feedback	Give information: statement
3. Student B	Teacher needs a microphone, as I can hardly understand what she is saying.	Technology	Report problem	Give information: statement
4. Student B	Didn't understand a word of what the student just said.	Technology	Report problem	Give information: statement
5. Student A	So far so good. Very helpful that the video stops when I try to make notes.	Technology	Provide feedback	Give information: statement

Table 2(b). Comparison of students' and academics' use of annotation notes – Video 2.

User type	Text in annotation note	Field	Purpose	Predominant speech function
1. Academic	Hi Everyone. Make sure you click and drag to turn the screen, so you can see what is going on behind you.	 Technology	Greeting Give advice	Greeting Demand action: command
2. Teacher	Hi All. I am the unit coordinator for this unit in Bentley and wrote the unit originally. This video I hope explains why we have set up so many of the aspects of the assessment the way we have. So if you wonder why, hopefully this will explain.	 Teaching & Learning	 Introduction Provide justification	Greeting Give information: statement
3. Teacher	Here I am asking you to use your metacognitive skills to start thinking like a teacher. You need to think of yourself as teacher and student, then how this unit and other role models practice.	Teaching & Learning	Provide guidance/ instruction	Give information: statement
4. Student	Find it really hard to read what is on the white board in the presentation.	Technology	Report problem	Give information: statement
5. Teacher	Sorry that it is hard to read. BUT it is in the PP notes so if you want to you can line them [up]. Problem statement: What are you researching? (in the 3rd person).	Technology Teaching & Learning	Apology Problem solution Provide guidance/ instruction	Give information: acknowledgement/ statement Demand information: question
6. Teacher	Think carefully about how you can increase the specificity of the project … Tigers: but (situational). For example Sumatran tigers over the last decade (time) — this helps your tutor determine that you have engaged and thought deeply about this topic.	Teaching & Learning	Provide guidance/ instruction	Demand action: command Give information: statement
7. Teacher	Please note the RQ are REALLY important and we do spend a lot of time helping you to refine them. It is difficult to get lots of feedback and corrections.	Teaching & Learning	Provide guidance/ instruction	Demand action: command Give information: statement

	BUT it is better to get help now. This is formative feedback and it is the most important form of assessment.			
8. Teacher	Why should you think about your Google search and the use of loaded words? What do we mean by that term? Can you give an example?	Teaching & Learning	Provide guidance/ instruction	Demand information: question

Implications and future directions

As the above analysis has shown, the information extracted from the data analytics allowed us to draw some conclusions about the ways in which students engaged with learning activities recorded from a 360-degree perspective. The results of our study resonate with the findings of Sarker's (2017) work on user engagement with mobile virtual reality narratives. His work shows that while well-designed audio-visual cues for directing users' attention in virtual environments are key for optimal user immersion and spatial presence, a lack of such cues can result in boredom, whereas excessive use of such cues may actually encourage "users to keep switching their attention from one element to the next in fear of missing out something important, eventually resulting in their frustration and stress" (Sarker, 2017, p. 437).

Approached from multimodal social semiotic perspective, the analysis further indicates that meaningful student engagement with educational content mediated through 360-degree videos requires a certain level of competency in negotiating and consolidating layers of different multimodal genres and context-based registers, e.g. instructional register, regulative register (e.g. see Bernstein 1990, 2000; Christie, 2005; Rose, 2014), and technology-oriented register (Tan, O'Halloran, & Wignell, 2016)—a task that only a fraction of students in our case study (28% for both videos) mastered successfully. The results have also made evident that teaching and learning with emergent technologies such as 360-degree video requires a deep understanding and appreciation of the complex ways in which multimodal resources (such as language, text, gesture, body posture, and use of space) work together to make meaning in different contexts, which has profound implications, not only for students, but also for academics. That is, apart from having to focus on the teaching of content, teachers need be aware of the meaning making potential of each semiotic mode or resource used in the creation of a lesson sequence to be recorded as a 360-degree panoramic video. It means that teachers not only need to be selective in choosing the right type of learning activity best suited to this mode of presentation, they also need to consider how to deliver educational content in these situations, and how it is to be captured, so that students can maximally benefit from the immersive panoramic perspective afforded by 360-degee videos, as compared to the unidirectional perspective offered by traditional video formats. In order to harness the full potential of emergent technologies such as 360-degree video, new methodological approaches to teaching and learning in such situations are thus much needed.

Of course, the present study is not without limitations. To begin with, the case study focuses only on students' interactions with two videos that were recorded for the same unit. Although the contextual situation for each recording was different, in all likelihood it involved the same cohort of students enrolled in the unit. Moreover, although the two videos that were selected for analysis had attracted the highest number of student views, the recorded learning activities and spatial contexts in which they were captured may not have constituted the most suitable scenarios for observing students' interactions with 360-degree video technology. Also, all inferences about student engagement are based solely on the information extracted from the data analytics which showed which areas in the videos students viewed based on their interactions with the 360-degree video viewer. As the study was not conducted in a controlled environment, observational data about what students actually looked at when viewing the videos was not available. In this respect, eye-tracking or gaze analysis (e.g. Blikstein & Worsley, 2016; Wang & Antonenko, 2017; Wang, Lin, Han & Spector, 2020) could prove useful for helping to form a more comprehensive understanding of what exactly attracts and captures (or distracts) student's attention over time. Similarly, advanced data visualisation models such as scanpath analyses, face clustering and attention maps (e.g. Sümer et al., 2018), as well as audio-visual and action recognition systems (e.g. Owens & Efros, 2018) could be useful for measuring users' viewing paths and attention in a more precise ways than currently possible. Wearable sensor data (e.g. Prieto, Sharma, Kidzinski, Rodríguez-Triana, & Dillenbourg, 2018) could provide additional information about users' kinetic efforts when interacting with 360-degree video player, and their ability to follow the unfolding activity on the screen. These and other computational visualisation tools and techniques could form the impetus for follow-up studies that investigate how viewers engage with 360-degree video content in teaching and learning situations.

Last but not least, as the 360-degree video platform and analytical approach presented in this paper are flexible and adaptable to other learning, teaching and training scenarios in other disciplines, it would be interesting to apply the same analytics techniques to other types of learning content mediated through 360-degree video technology, such as museum exhibits, outdoor activities, sporting events, etc. Indeed, a follow-up study that applies the 360-degree video technology developed it this project in the context of a pre-school environment is already underway.

In this sense, the present study can also be a first step toward the study and use of HMD VR technologies as effective pedagogical tools, with a view to enhancing student experiences with immersive technologies, and ultimately leading to a better understanding of which educational activities and learning contexts are best suited to immersive video formats, particularly in subjects where it is important for learners to "mentally adopt the spatial perspective of others [to] understand the world from their point of view (Kessler & Thomson, 2010, p. 72).

Acknowledgments

This work extends findings from the project "Encouraging and mapping student engagement through 360-degree video annotation and data analytics" with Lead Chief Investigator Professor Kay O'Halloran, funded by the Australian Technology Network (ATN) (for further information, see: http://online360video.education/).

Declaration of interest statement

No potential conflict of interest was reported by the authors.

References

Alzahrani, A., Gardner, M., Callaghan, V., & Alrashidi, M. (2015, July). Towards measuring learning effectiveness considering presence, engagement and immersion in a mixed and augmented reality learning environment. In *Intelligent Environments (Workshops)* (pp. 252-264).

Argyriou, L., Economou, D., & Bouki, V. (2017). 360 degree interactive video application for cultural heritage education. In D. Beck, C. Allison, L. Morgado, J. Pirker, F. Khosmood, J. Richter, & C. Gütl (Eds.), iLRN 2017: Coimbra workshop, long and short paper, and poster proceedings from the third immersive learning research network conference (pp. 297–304). Verlag der Technischen Universität Graz.

Andrade, A., Delandshere, G., & Danish, J. A. (2016). Using multimodal learning analytics to model student behavior: A systematic analysis of epistemological framing. *Journal of Learning Analytics*, *3*(2), 282-306.

Assilmia, F., Yun, S. P., Okawa, K., & Kunze, K. (2017). IN360: A 360 degree-video platform to change students preconceived notions on their career. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2359–2365). New York: ACM.

Bernstein, B. (1990). *Class, codes and control: The structuring of pedagogic discourse* (Vol. 4). London: Routledge.

Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique*. London: Taylor & Francis.

Blikstad-Balas, M. (2017). Key challenges of using video when investigating social practices in education: Contextualization, magnification, and representation. *International Journal of Research & Method in Education*, *40*(5), 511–523.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238.

Boling, E. C., Hough, M., Krinsky, H., Saleem, H., & Stevens, M. (2012). Cutting the distance in distance education: Perspectives on what promotes positive, online learning experiences. *The Internet and Higher Education*, *15*(2), 118–126.

Chittleborough, G., Cripps Clark, J., & Chandler, P. (2015). The pedagogy of using video to develop reflective practice in learning to teach science. In E. Ortlieb, L. E. Shanahan, & M. B. McVee (Eds.), *Video research in disciplinary literacies* (pp. 95–115). Emerald Group Publishing Limited.

Ciekanski, M., & Chanier, T. (2008). Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment. *ReCALL*, *20*(2), 162–182.

Christie, F. (2005) *Classroom discourse analysis: A functional perspective*. London & New York: Continuum.

Crook, A., Mauchline, A., Maw, S., Lawson, C., Drinkwater, R., Lundqvist, K., ... Park, J. (2012). The use of video technology for providing feedback to students: Can it enhance the feedback experience for staff and students?. *Computers & Education*, *58*(1), 386–396.

Davidsen, J., & Vanderlinde, R. (2014). Researchers and teachers learning together and from each other using video-based multimodal analysis. *British Journal of Educational Technology*, *45*(3), 451–460.

Eggins, S. (2005). *An introduction to systemic functional linguistics* (2nd ed.). London: Pinter Publishers.

Elmezeny, A., Edenhofer, N., & Wimmer, J. (2018). Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion. *Journal for Virtual Worlds Research, 11*(1), 1–13.

Gill, P., & Remedios, R. (2013). How should researchers in Education operationalise on-task behaviours?. *Cambridge Journal of Education*, 43(2), 199–222.

Georgiou, Y., & Kyza, E. A. (2017, October). A design-based approach to augmented reality location-based activities: Investigating immersion in relation to student learning. In *Proceedings of the 16th World Conference on Mobile and Contextual Learning* (pp. 1-8).

Goodwin, C., & Duranti, A. (1992). Rethinking context: An introduction. In C. Goodwin & A. Duranti (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 191–227). Cambridge: CUP.

Gregory, S., Lee, M. J., Dalgarno, B., & Tynan, B. (2016). *Learning in virtual worlds: Research and applications*. Athabasca University Press.

Hackling, M. W., Murcia, K. J., Ibrahim-Didi, K., & Hill, S. M. (2014). Methods for multimodal analysis and representation of teaching-learning interactions in primary science lessons captured on video. *Proceedings of Conference of the European Science Education Research Association* (pp. 1–8). University of Cyprus, Nicosia, Cyprus. European Science Education Research.

Hales, P., & Kalyvaki, M. (2017). Immersive learning experiences in teacher education. In *Society for Information Technology & Teacher Education International Conference* (pp. 1586–1587). Waynesville, NC: Association for the Advancement of Computing in Education.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. London & New York: Routledge.

Healion, D., Russell, S., Cukurova, M., & Spikol, D. (2017, March). Tracing physical movement during practice-based learning through multimodal learning analytics. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 588–589). ACM.

Hung, S.-T. A. (2016). Enhancing feedback provision through multimodal video technology. *Computers & Education, 98*, 90–101.

Hung, I. C., & Chen, N. S. (2018). Embodied interactive video lectures for improving learning comprehension and retention. *Computers & Education*, 117, 116–131.

Jewitt, C. (2003). Re-thinking assessment: Multimodality, literacy and computer-mediated learning. *Assessment in education: Principles, policy & practice*, *10*(1), 83–102.

Jewitt, C. (2012). *An introduction to using video for research*. National Centre for Research Methods Working Paper 03/12.

Jewitt, C., Kress, G., Ogborn, J., & Tsatsarelis, C. (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, *53*(1), 5–18.

Kessler, K. & Thomson L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72–88.

Leijon, M. (2016). Space as designs for and in learning: Investigating the interplay between space, interaction and learning sequences in higher education. *Visual Communication*, *15*(1), 93–124.

Lim, F. V., O'Halloran, K. L., & Podlasov, A. (2012). Spatial pedagogy: Mapping meanings in the use of classroom space. *Cambridge Journal of Education*, *42*(2), 235–251.

Major, L., & Watson, S. (2018). Using video to support in-service teacher professional development: The state of the field, limitations and possibilities. *Technology, Pedagogy and Education*, *27*(1), 49-68.

Martin, J. R. (2002). Meaning beyond the clause: SFL perspectives. *Annual Review of Applied Linguistics*, *22*, 52–74.

Martin, J. R., & Rose, D. (2007). W*orking with discourse: meaning beyond the clause* (2nd ed). London: Continuum.

Martin, J. R., & White, P. R.R. (2005). *The language of evaluation, appraisal in English*. London & New York: Palgrave Macmillan.

McGowan, A., & Hanna, P. (2015, November). How video lecture capture affects student engagement in a higher education computer programming course: A study of attendance, video viewing behaviours and student attitude. In *eChallenges e-2015 Conference,* 2015 (pp. 1–8). IEEE.

Morell, T. (2015). International conference paper presentations: A multimodal analysis to determine effectiveness. *English for Specific Purposes*, *37*, 137–150.

Morell, T. (2017). Multimodal competence and effective interactive lecturing. *System*. Advance online publication. doi.org/10.1016/j.system.2017.12.006

O'Halloran, K. L. (2011). The semantic hyperspace: Accumulating mathematical knowledge across semiotic resources and modes. In F. Christie & K. Maton (eds), *Disciplinarity: Functional linguistic and sociological perspectives*. London: Continuum, 217–236.

O'Halloran, K. L. (2015). The language of learning mathematics: A multimodal perspective. *The Journal of Mathematical Behaviour, 40* Part A, 63–74.

O'Halloran, K. L., Tan, S., Wiebrands, M., Sheffield, R., Wignell, P., & Turner, P. (2018). The multimodal classroom in the digital age: The use of 360 degree videos for online learning. In H. de Silva Joyce & S. Feez (Eds.), *Multimodality across classrooms: Learning about and through different modalities* (pp. 84–102). London & New York: Routledge.

Owens, A., & Efros, A. A. (2018). Audio-visual scene analysis with self-supervised multisensory features. *arXiv preprint arXiv:1804.03641*.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, *34*(2), 193–203.

Rose, D. (2014). Analysing pedagogic discourse: An approach from genre and register. *Functional Linguistics*, *1*(11), 1–32.

Ryan, J., Scott, A., & Walsh, M. (2010). Pedagogy in the multimodal classroom: An analysis of the challenges and opportunities for teachers. *Teachers and Teaching: Theory and Practice*, *16*(4), 477–489.

Salselas, I., & Penha, R. (2019, September). The role of sound in inducing storytelling in immersive environments. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound on ZZZ* (pp. 191-198).

Sakr, M., Jewitt, C., & Price, S. (2016). Mobile experiences of historical place: A multimodal analysis of emotional engagement. *Journal of the Learning Sciences*, *25*(1), 51–92.

Sarker, B. (2017, July). Decoding the user experience in mobile virtual reality narratives. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 437-452).

Sauli, F., Cattaneo, A., & van der Meij, H. (2018). Hypervideo for educational purposes: A literature review on a multifaceted technological tool. *Technology, Pedagogy and Education*, *27*(1), 115–134.

Sheikh, A., Brown, A., Watson, Z., & Evans, M. (2016). Directing attention in 360-degree video. IBC 2016 Conference. DOI: 10.1049/ibc.2016.0029.

Shernoff, D. J. (2013). *Optimal learning environments to promote student engagement*. New York, NY:

Smith, C., King, B., & Gonzalez, D. (2016). Using multimodal learning analytics to identify patterns of interactions in a body-based mathematics activity. *Journal of Interactive Learning Research*, *27*(4), 355–379.

Sümer, O., Goldberg, P., Stürmer, K., Seidel, T., Gerjets, P., Trautwein, U., & Kasneci, E. (2018). Teachers' perception in the classroom. *arXiv preprint arXiv:1805.08897*.

Tan, S., O'Halloran, K. L., & Wignell, P. (2016). Multimodal research: Addressing the complexity of multimodal environments and the challenges for CALL. ReCALL, *28*(3), 253–273.

Twiner, A., Coffin, C., Littleton, K., & Whitelock, D. (2010). Multimodality, orchestration and participation in the context of classroom use of the interactive whiteboard: A discussion. *Technology, Pedagogy and Education*, *19*(2), 211–223.

van Es, E. A., Cashen, M., Barnhart, T., & Auger, A. (2017). Learning to notice mathematics instruction: Using video to develop preservice teachers' vision of ambitious pedagogy. *Cognition and Instruction, 35*(3), 165–187.

van Es, E. A., Stockero, S. L., Sherin, M. G., Van Zoest, L. R., & Dyer, E. (2015). Making the most of teacher self-captured video. *Mathematics Teacher Educator, 4*(1), 6–19.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89.

Wang, X., Lin, L., Han, M., & Spector, J. M. (2020). Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos. *Computers in Human Behavior*, *107*, 106279.

Zappavigna, M., & Zhao, S. (2017). Selfies in 'mommyblogging': An emerging visual genre. *Discourse, Context & Media*, *20*, 239–247.