Education Tech Research Dev (2023) 71:881–899
https://doi.org/10.1007/s11423-023-10215-2

RESEARCH ARTICLE

Exploring the relationships between teacher noticing, ambisonic audio, and variance in focus when viewing 360 video

Richard E. Ferdig[1] · **Karl W. Kosko**[1] · **Enrico Gandolfi**[1]

Accepted: 3 March 2023 / Published online: 4 April 2023
© Association for Educational Communications and Technology 2023

Abstract
A growing body of research has supported the implementation of innovative and immersive video for teaching and learning across the lifespan. Immersive video, delivered through eXtended Reality (XR) tools like 360 video, provides users with new ways to see real or created environments. Unfortunately, most of the existing research has highlighted immersive video without accompanying immersive audio. This use of monophonic audio can create a disconnect for viewers as they experience close to real world video with sounds that do not match a real-world environment. The purpose of this study was to respond to this gap in the literature by exploring the use of ambisonic audio and its impact on pre-service teacher noticing and variability of viewing focus when watching 360 video. Data were collected from undergraduate teacher education students who participated in a self-paced online activity that included watching 360 videos and responding to a questionnaire. A convergent mixed methods design was employed to compare participants' professional noticing and observed viewing behavior in the context of ambisonic and monophonic audio. Results showed that users in ambisonic audio conditions in 360 video environments were more likely to have higher focus. Moreover, for users who had specific professional knowledge, monophonic audio with immersive video had a negative impact on their variance in focus. The paper concludes with recommendations for future research on the use of audio in virtual and augmented reality environments.

Keywords Ambisonic audio · Extended reality · Teacher education · Teaching noticing

Introduction

Video representation is used widely for teaching and learning across the lifespan (Poquet et al, 2018). Its applications range from teaching surgery preparation (Mota et al., 2018) to improving attention in kindergarten students (Sihotang et al., 2021). As technologies have continued to evolve, so have adaptations for educating with video. For instance, researchers

✉ Richard E. Ferdig
rferdig@gmail.com

[1] Kent State University, 800 E Summit St, Kent, OH 44240, USA

have explored the ways in which video can be used for teaching and learning with and through immersive tools like augmented reality (Yip et al., 2019), virtual reality (Jong et al., 2020), and 360 cameras (Kosko et al., 2021a).

Less work has focused on the relationship between audio and immersive video for educational outcomes. Audio arguably received significant attention in the early 1990s with dual coding theory (Clark & Paivio, 1991; Paivio, 1990) and research on multimodal learning (Mayer & Sims, 1994); both suggested significant gains could occur from representations that included more than one modality. Moreover, developing a research basis for educational audio was significantly impacted by the advent of audio-based technologies like podcasts (e.g., Hargett, 2018; Prakash et al., 2017). Finally, early work has also begun to explore the uses of immersive audio in education (e.g., Ferdig et al., 2020) and beyond (e.g., Rumsey, 2020; Yao, 2017). Much of the aforementioned work, however, is demonstrative, theoretical, or technical, with few reports empirically describing the relationships between immersive video, audio, and learning.

More recently, researchers have begun to explore differences in the use of ambisonic vs. monophonic audio. Monophonic audio, at least in the context of immersive environments, refers to sound that has the same volume and directionality, regardless of where the user looks or walks in the virtual environment. Comparatively, ambisonic audio presents sounds located in three-dimensional space (Malham & Myatt, 1995). Such sound is said to have directionality, increasing in volume or intensity based on the gaze or location of the user in the immersive experience. Researchers have offered early evidence that such audio delivery can increase focus and attention (Ferdig et al., 2020), support the representation of the content (Fela & Zacharov, 2020; Ferdig & Kosko, 2020), and facilitate awareness of the user in the virtual space (Gandolfi et al., under review). While promising, little is known about teaching with ambisonic audio in the context of the increased use of 360 video in teacher education.

This study set out to address this dearth of research by comparing preservice teachers (PSTs) as they engaged with 360 video in either monophonic or ambisonic audio conditions. The mixed-methods study examined the noticings and variability in viewing focus of both elementary education and education majors in other disciplines (e.g., middle and high school) who watched a recorded elementary grades math lesson. There were two research questions guiding this study:

(1) What impact, if any, does ambisonic audio have on the variability of viewing focus of PSTs as they watch a 360 immersive video of a math lesson?
(2) What interaction, if any, does professional knowledge have on the noticings and variability of viewing focus, compared across ambisonic and monophonic conditions, as PSTs watch a 360 immersive video of a math lesson?

Literature review

Professional noticing

Key in the education of future teachers is the development of professional noticing (AMTE, 2017). Teachers' professional noticing involves the three interrelated skills of *attending* to key aspects of pedagogy, *interpreting* these events through the lens of professional knowledge, and *deciding* how to respond or act next as the teacher (Jacobs et al., 2010; van Es et al., 2017). Initially, novice teachers tend to focus on more generic aspects of a classroom, such as students' behavior, classroom management, etc. Over time, these same teachers can learn to focus on students' procedures and then to interpret those procedures using pedagogical content knowledge (Barnhart & van Es, 2015; Schack et al., 2013). In describing such transitions, van Es et al. (2017) analyzed how a group of elementary preservice teachers (PSTs) used video to articulate aspects of mathematics teaching and learning. They found that as PSTs improved the sophistication of their noticing, there was a corresponding progression in the type of language used to describe such pedagogy (also see Jacobs et al., 2010; Schack et al., 2013). Stated differently, as teachers learn to attend to more specific facets of students' reasoning, they use more specific and technical language in their descriptions.

Scholars studying professional noticing in mathematics classrooms have typically identified variations of three characteristics for facilitating it: focusing on students' reasoning about the content (Jacobs et al., 2010; van Es et al., 2017); focusing on and decomposing particular events within a representation of practice (Brunvand & Fishman, 2007; Schack et al., 2013; Teuscher et al., 2017); and increasing the quality of the representation used for noticing (Kosko et al., 2021b; Seidel et al., 2011). Specifically, by prompting PSTs to focus on particular aspects of children's mathematics, PSTs begin to decompose, or break down, observable student actions to articulate specific content-related details (Jacobs et al., 2010; van Es et al., 2017). However, the interaction between what and how PSTs decompose practice is interrelated with the nature of how professional practice is represented (Kosko et al., 2021a). For example, Kosko et al. (2021b) observed that PSTs who used 360 video to view the same recorded scenario as peers using standard video tended to describe more specific student actions and with more specificity toward the mathematics that recorded students were engaged. This corresponds with findings of others who observed that 360 video allowed for PSTs to focus more on students' mathematical thinking in their reflective noticing (Buchbinder et al., 2021; Weston & Amador, 2021).

More immersive representations of practice (i.e., 360 video and VR-based scenarios) have been advocated for facilitating PSTs' reflection and professional noticing because they are "able to re-experience their teaching, emplaced within its space and time…and with agency to select where and with what to engage" (Walshe & Driver, 2019, p. 103). Kosko et al., (2021a, 2021b) suggested that some of the benefits associated with more immersive representations is due to such mediums' perceptual capacity. *Perceptual capacity* is the capacity of a particular representation to re-present aspects of an individual's actual or potential embodied experience. In other words, professional knowledge is embodied in teachers' experiences, with professional noticing being one expression of such knowledge (Dessus et al., 2016; Grub et al., 2020; Huang et al., 2021; Kosko et al., 2021a). Eye-tracking research with teachers have found novices tend to have more erratic gaze behavior, as they appear to focus on multiple students at the same time (Dessus et al., 2016; Grub et al., 2020). By contrast, more experienced teachers may look at more students but spend most of the time focusing on a select subsample of students (Huang et al., 2021). Such research with eye-tracking corresponds with findings from scholars examining teachers' changes in their field of view (FOV) in 360 videos (Kosko et al., 2021a) and animated VR environments (Huang et al., 2021). Specifically, Gandolfi et al. (2021b) observed that PSTs who had more variance in their FOV had less sophisticated professional noticing than their peers with lower variance in their FOV.

Given evidence from prior research regarding the embodied nature of teachers' professional actions, such as noticing, and the role that perceptual capacity plays in representing practice, there is a need to better understand how various facets of immersive technology affect teachers' professional noticing and the pragmatic magnitude of such effects. For example, PSTs appear to notice more student content-related actions when watching a 360 video on a flat screen device (i.e., laptop) than from a standard video of the same class scenario, but also notice more student actions watching the same 360 video with a VR headset than on a flat-screen device (Ferdig & Kosko, 2020; Kosko et al., 2021b). Yet, beyond a handful of studies, there is little scholarship regarding the pragmatic effect a representation's capacity to re-present practice (i.e., perceptual capacity) and the degree to which teachers attend to practice meaningfully.

The audio disconnect

Video and audio have not always been synchronized in their delivery for educational purposes or otherwise. Consider, for instance, silent films (Klepper, 2005) or even the more recent research on educational, audio podcasts (Hargett, 2018). However, early work by Clark and Paivio (1991) and Mayer and Sims (1994) lent credence to the value of having both audio and visual representations for increased knowledge consumption.

Video creation, production, and delivery have gone through significant changes to improve viewability; related studies have addressed such turning points (e.g., Cranley et al., 2006). Audio has also gone through related advancements, increasing factors like quality, compression, and streaming (e.g., Reddy & Vijayarajan, 2020). It is interesting to note, however, that the use of new technologies for delivering educational video have more often been focused on the delivery of innovative and immersive video—the sight—without addressing the need for accompanying audio—the sound. For instance, numerous studies have addressed the implementation of *eXtended Reality* (XR) video—i.e., videos that aim to expand the sensorial involvement of their viewers—for learning (e.g., Chen, 2020; Dong & Li, 2021; Ulrich et al., 2021). These studies should be valued for the exploration of innovative viewing experiences through augmented reality (AR), virtual reality (VR), and 360 video. But the inclusion of XR video without accompanying XR audio could create feelings of cognitive dissonance (Festinger, 1962; see also the relationship between sound and cognitive dissonance in Masataka & Perlovsky, 2012).

In the example of 360 video, users are presented with an opportunity to look in any direction. Such scenarios often promote feelings of presence or a sense of being there (Gandolfi et al., 2021a), particularly when compared to the use of standard, unidirectional video (Gold & Windscheid, 2020). However, sound in 360 video experiences is often delivered monophonically. Said differently, users have the ability to look around as they would in real life; but, unlike real life, sound is often delivered to them in every direction. It is nearly impossible for viewers to tell where sound is coming from or what actor in the scene produced a given sound without looking directly at the speaker.

This combination of immersive video with monophonic sound could create feelings of dissonance and a reduced sense of presence. More importantly, at least for the use of teaching with XR videos, users in monophonic conditions have more variability in where they look during the video (Ferdig et al., 2020). They spend significant amounts of time looking around to place the audio they are hearing; in doing so, they lose the ability to notice key events in the video tied to the content being offered.

Some researchers have attempted to address this imbalance with the introduction of ambisonic audio (Gupta et al., 2017; Li et al., 2018; Narbutt et al., 2018). Ambisonic audio can broadly be defined as "a two-part technological solution to the problems of encoding sound directions (and amplitudes) and reproducing them over practical loudspeaker systems so that listeners can perceive sounds located in three-dimensional space" (Malham & Myatt, 1995, p. 62). Figure 1 contains a representation of a user watching a 360 video and being able to hear sound directionally.

Immersive audio has been associated with noteworthy benefits with a broad range of applications. For instance, Rumsey (2020) highlighted how it can make mediated environments easier to understand and control, particularly for users who are not experienced listeners and who have different hearing preferences. Yao (2017) found that 3D audio can also address issues related to frequency hearing impairment, localization error, and localization blur in virtual reality settings. In addition, Brandmeyer et al. (2021) analyzed how ambisonic audio was able to empower users' reaction times in gaming by allowing them to read matches more effectively. Finally, Chen et al. (2020) developed and tested a navigation system for VR environments that relied on immersive audio inputs; they found that 3D sound improved the ability to visualize and move across digital scenarios and embodied cognition processes (see also Gandolfi & Clements, 2019).

Drawing on this audio work, educational researchers have provided early evidence that ambisonic audio can positively impact immersive video delivery in three key ways. First, it can lead to increased noticing and visual focus (Ferdig et al., 2020). Second, it supports increased "perceptual capacity" (Ferdig & Kosko, 2020, p. 856), or the ability of the technology to represent content that can be obtained by the user when compared with standard video. Third, it facilitates viewers' awareness of the environment recorded in terms of reported focus of attention, especially when camera placement is defective (Gandolfi et al., under review).

Even with these promising early results, little is known about teaching with ambisonic audio to accompany XR video. Promising technical work continues to make ambisonic audio available (Narbutt et al., 2018), realistic (Fela & Zacharov, 2020), and even automated (Yang et al., 2020). However, the field lacks a large body of empirical literature related to ambisonic audio and user performance. This study set out to address this gap by exploring the impact of ambisonic audio on the learning performance of preservice teachers (PSTs) in a large university in the northeast United States.

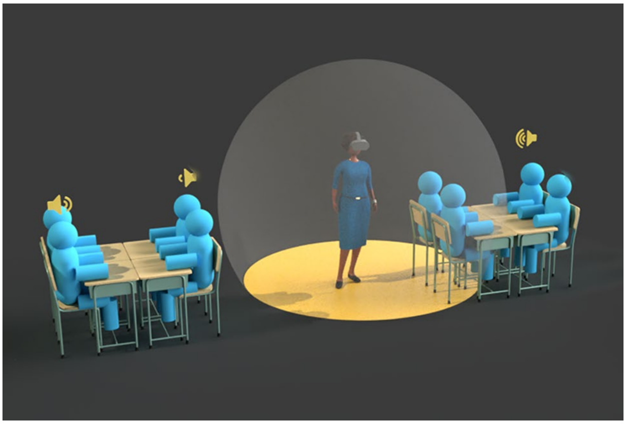

Fig. 1 Illustration of ambisonic audio in VR

Methods

The study relied on a self-paced online activity that included watching 360 videos and completing a related questionnaire. Data collection was approved and monitored by the authors' university Institutional Review Board.

Sample

Participants included 21 preservice teachers (PSTs) enrolled in a teacher licensure program (PreK-grade 3; ages 3–9 years) at a large Midwestern U.S. university. The majority of participants self-identified as female (76.2%) and white (95.2%). Specifically, one participant identified as a biracial female, five as white males, and 15 as white females. Participating PSTs were preparing to teach in a wide range of domains, with 19% focusing on early childhood ($n=4$), 14.3% focusing on middle or secondary mathematics ($n=3$), and the remainder focusing on some other area of education. Thus, 33.3% of participants were seeking licensure in an area that related to the video in some fashion (i.e., teaching upper elementary mathematics). When asked about their perceived technological savviness and prior experience with 360 video, participants viewed themselves as more technologically savvy than not ($M=6.53$, $SD=1.71$ on a 10-point scale), with 75% having viewed a 360 video prior to the study.

Procedure

This study was among the options for fulfilling an undergraduate research credit requirement; it had to be completed online due, in part, to study requirements during COVID-19. The procedure was entirely self-paced. Students began with an online questionnaire that included basic demographic data like gender, perceived technological savviness (with a ten-point Likert scale ranging from not all tech savvy to extremely tech savvy), and their specific education major like mathematics or literacy education. The survey also addressed their familiarity with 360 video. All participants were then instructed to watch a 3-min tutorial video on how to view and navigate 360 videos.

PSTs were then randomly assigned to one of two groups. The first group, the monophonic (control) group, received a 5-min, 49-s, 360 video about an elementary classroom mathematics lesson. The teacher in the video lesson introduces equivalent fractions to her fourth-grade students. PSTs watched the video on their desktop or laptop with headphones; they used their mouse or laptop tracking device to look around the room. The audio related to the video track was monophonic; in other words, regardless of where they looked in the video, they heard the sounds with the same intensity, volume, and directionality. The second group—the ambisonic (experimental) group—watched the same 360 video on their desktop or laptop with headphones. However, their audio recording was ambisonic. Sound came from the direction of the source (e.g., a student, the teacher) in the video (see Fig. 1).

After watching the first video, both groups were asked to describe pivotal moments they saw in the video. A pivotal moment was considered something they felt was important and related to teaching and learning they observed in the classroom. They were then asked to watch the video a second time (within the same audio condition), during which they were asked to turn on screen recording so that researchers could see what they saw. After the second video, both groups were asked to select one of the pivotal moments they noticed and describe why they labeled it as pivotal or important.

The final step was for participants to complete a survey that consisted of two components. The first part measured perceived presence through the *Extended Reality Presence Scale*, or XRPS (Gandolfi et al., 2021b). The second part of the survey was a map of the classroom. Participants were asked to pick ten places they focused on while watching the video. All 21 subjects completed the pre- and post-survey instruments; all the students also completed and uploaded their screen recordings of their watched XR sessions.

Analysis

This study employed a convergent mixed methods design to compare participants' professional noticing and observed viewing behavior in the context of ambisonic and monophonic audio. Mixed methods incorporate qualitative and quantitative analyses, merging the results and findings to form a more cohesive image of the phenomenon of study. In a convergent design, qualitative and quantitative analysis is conducted at the same stage of study and may often inform each other bidirectionally (Creswell & Plano Clark, 2018). In the present study, PSTs' written descriptions of what they noticed (or attended to) in the 360 video were examined qualitatively. Video recordings were coded quantitatively to identify portions of the classroom and the amount of time (and which times) participants included those areas in their field of view (FOV). The data were then merged to examine how qualitative themes and quantitative measures of variance in where participants focused their FOV related. Specifically, results of the quantitative and qualitative analysis allowed for supplemental examination of segments of the 360 video where participants appeared to focus. These analyses are described in more detail below.

Qualitative analysis and findings

Analysis of participants' written noticing was examined using *Systemic Functional Linguistics* (SFL). SFL focuses on how grammar functions to convey meaning (Halliday & Matthiessen, 2004). For the present study, the focus was on the grammatical system of reference, which "creates cohesion by creating links between elements" (p. 534) within conveyed language. For example, one participant describing what they attended to in the 360 video noted that:

> The teacher brought up students trying to divide 5/6 to get an equivalent fraction. They realized they would not be able to divide since 5 is a prime number.

Here, the key references to mathematics at-hand (i.e., the content children were learning in the video) are underlined to illustrate how, initially, reference is conveyed. First the participant references "dividing 5/6" as a means to "get an equivalent fraction." Such a sequencing of references is nontrivial as it forms what Halliday and Matthiessen (2004) describe as an information unit, which "is the tension between what is already known or predictable and what is new or unpredictable" (p. 89). The reference chain constructed in this brief excerpt extends to incorporate an additional information unit noting that 5 is prime. Considering these individual references as a reference chain that goes to construct a particular experience, the participant is attending to mathematics, but also the procedures children used to find a particular equivalent fraction and a rationale for why they did this. A goal of this study was to examine how participants conveyed meaning regarding children's mathematics, as this has been identified as a key area for improving teacher education and teacher noticing (Jacobs et al., 2010).

Findings corresponded with those from prior research on teachers' mathematical noticing (Jacobs et al., 2010; van Es et al., 2017) with participants referencing general aspects of the classroom with no reference or reference chains regarding mathematics, participants focusing on children's procedures for the mathematics, and/or participants referencing conceptual aspects of mathematics such as children's rationales and reasoning beyond procedures. The example in the preceding paragraph illustrates attending to procedures but also going beyond to focus on children's rationales and reasoning about such procedures. The other two themes (no math & procedural) are exemplified in Table 1. Whereas the non-math specific example does reference students' answers, such referencing could point to any content area and is not, in and of itself, referencing anything particularly mathematical. By contrast, the procedural excerpt references a primary-grades mathematical tool used by the teacher in this lesson for mathematics (fraction strips), the specific fraction children were working on, and how the children had to find an alternate strategy to solve the problem at-hand. Across both examples, the reference chains exemplify the *focus* of what participants attended, and not merely whether something was mentioned in passing. The three themes were relatively evenly distributed across monophonic (*non-math* = 4; *procedural* = 3; *conceptual* = 3) and ambisonic (*non-math* = 6; *procedural* = 2; *conceptual* = 3) conditions.

Following an initial round of analysis of reference chains, two coders carried out a second round of analysis to categorize participants' writings. The Kappa coefficient was used as an indicator for reliability and was found to be 0.85, indicating near perfect agreement (Landis & Kotch, 1977). Codings were reconciled and later merged with quantitative data for analysis.

Quantitative analysis and results

Participants' variability in focus in their FOV was measured using the unalikeability statistic. *Unalikeability* (U_2) is a nonparametric measure for nominal variables corresponding to variance associated with a parametric mean (Kader & Perry, 2017). In the present sample, participants' count data for each region (i.e., seconds per region) was used to estimate how alike or unalike the frequencies were. Figure 2 illustrates the regions of the classroom, as well as the camera placement. This resulted in the U_2 statistic where participants with lower U_2 statistics had more unalike frequencies (i.e., more seconds spent on specific regions) and participants with higher U_2 statistics had more alike frequencies (i.e., more similar amount of time spent across different regions). U_2 statistics were calculated for all participants and are presented as aggregated per comparison group in Table 2. Seconds

Table 1 Participants' excerpts illustrating the primary themes

Non-math specific	Math specific (procedural)
I like when the students **shared** their different answers and the class **collaborated** together to try and come to a conclusion When the students **shared** what answers they got and **worked** together, talking through the problem. Sometimes it **is** helpful for students to **discuss** their thought process and work with their peers to come to a conclusion	The teacher challenged her students to **use** the fraction strips for the second fraction (3/8). I noticed the students closest to the 360 camera quickly realized that the second problem (3/8) *could not* be **demonstrated** using the fraction strips. It is important that we as teachers monitor our students while they work and **share** students' realizations when they come to them

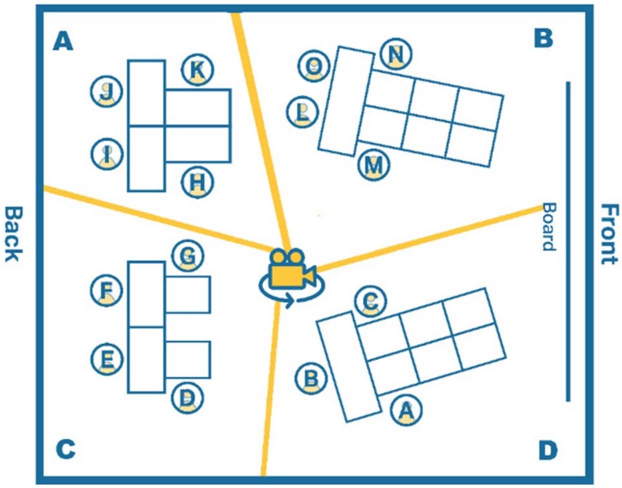

Fig. 2 Classroom map divided into four viewing regions

Table 2 Distribution of seconds viewed by region and corresponding unalikeability statistic

	n	Region A	Region B	Region C	Region D	Indeterminant	U_2
Monophonic$_{NonMath}$	4	102	537	188	514	5	.650
		76.28	*436.45*	*177.30*	*637.56*	*18.42*	
Monophonic$_{Procedural}$	3	84	252	135	570	12	.710
		59.67	*341.44*	*138.70*	*498.77*	*14.41*	
Monophonic$_{Conceptual}$	3	91	308	154	419	27	.663
		56.61	*323.93*	*131.59*	*473.20*	*13.67*	
Ambisonic$_{NonMath}$	6	64	620	272	1108	42	.570
		119.34	*682.89*	*277.40*	*997.55*	*28.82*	
Ambisonic$_{Procedural}$	2	34	244	63	358	3	.596
		39.78	*227.63*	*92.47*	*332.52*	*9.61*	
Ambisonic$_{Conceptual}$	3	35	385	141	458	10	.423
		58.31	*333.66*	*135.54*	*487.41*	*14.08*	

Seconds assigned as indeterminant were not included in calculating U_2 statistics. Italicized numbers are seconds expected by chance with unitalicized numbers being observed seconds

assigned to the indeterminant category were not included since this category represented uncertainty on the part of coders and could potentially inflate the U_2 statistic.

The difference in U_2 statistics between the monophonic and ambisonic conditions was compared using the Mann–Whitney test, which is a non-parametric corollary to the independent samples t-test (Siegel & Castellan, 1988). The Mann–Whitney test was ideal for this comparison given the size of comparison groups (monophonic = 10; ambisonic = 11). The same statistical analysis was used to examine differences between monophonic and ambisonic regarding the level of detail provided in participants' written noticing (not noticed, procedural math noticed, conceptual math noticed).

Following comparison of main effects, comparison groups were created by merging the results of the qualitative analysis of written noticing with the category participants were assigned (monophonic vs. ambisonic). The Kruskal–Wallis one-way analysis of variance was used to examine median differences in unalikeability scores across the six groups comparing their audio condition and their noticing of instructional math strategies ($Monophonic_{Not\_Noticed}$, $Monophonic_{Math\_Noticed}$, $Ambisonic_{Not\_Noticed}$, $Ambisonic_{Math\_Noticed}$). The Kruskal–Wallis test is a nonparametric variation of ANOVA that uses sum of ranks to compare average ranks across multiple groups (Siegel & Castellan, 1988). It was ideal for the present study, given the relatively small sample size ($n = 21$). In tandem with the Kruskal–Wallis analysis of variance, Dunn's (1964) test for multiple comparisons was implemented as a post hoc analysis. Essentially, Dunn's test is a nonparametric form of post hoc that uses rank sums. Given that multiple comparisons introduce the risk of a Type I error (false positive), Simes (1986) adjustment procedure for p-values was used. Simes' (1986) approach is an improvement over the common Bonferroni approach with the additional benefit of reducing the risk of a Type II error (false negative) that the Bonferroni adjustment can introduce.

Results

The unalikeability scores of participants in the monophonic condition were found to be higher on average than for participants in the ambisonic condition, and this difference was statistically significant ($U = 10.0$, $p < 0.001$). However, when comparing for differences in specificity of noticing, no statistically significant difference was observed ($U = 48.5$, $p = 0.674$). This result was interesting given the fact that teachers' specificity of noticing is associated with the variance in where they attend when viewing 360 video. To further explore this finding, subgroups described previously (see Table 2) were used to explore differences in unalikeability scores that may be associated with both the audio condition and the specificity participants attended to children's mathematics.

Comparison of unalikeability scores across the six groups was found to be statistically significant ($KW(df = 5) = 12.237$, $p = 0.032$). Dunn's (1964) post hoc analysis, with Simes (1986) p-value adjustment, indicated statistically significant differences in pairwise comparisons between $Ambisonic_{Non\text{-}Math}$ and $Monophonic_{Procedural}$, as well as $Ambisonic_{Conceptual}$ and $Monophonic_{Conceptual}$. All other pairwise comparisons were found to be not statistically significant from chance. Table 3 highlights nine of the 15 pairwise comparisons (monophonic vs ambisonic) to aid the reader in noting the statistically significant relationships.

Figure 3 provides a graphical illustration of the medians and associated variance for participants' U_2 scores across conditions. Important in considering both the statistical and graphically represented differences is the relatively low sample size in the current study. Despite this lower sample size, participants in the monophonic condition demonstrated higher variance in their FOV than participants in the ambisonic condition, with a similar trend in terms of their noticing behaviors. It is also worth noting in these trends that participants attending to mathematical procedures in the monophonic condition had statistically significant higher variance than participants in the ambisonic condition who attended to generic (non-math) classroom events and those who attended to students' conceptual mathematics.

Table 3 Selected post hoc results from Dunn's pairwise comparisons with Simes adjusted p-values

	$\text{Monophonic}_{\text{Non-math}}$	$\text{Monophonic}_{\text{Procedural}}$	$\text{Monophonic}_{\text{Conceptual}}$
Ambisonic $_{\text{Non-Math}}$	$z=6.083$	$z=11.917$	$z=6.917$
	S.E. = 3.987	S.E. = 4.367	S.E. = 4.367
	$p=.272$	$p=.045$	$p=.272$
Ambisonic $_{\text{Procedural}}$	$z=5.250$	$z=11.083$	$z=6.083$
	S.E. = 5.349	S.E. = 5.638	S.E. = 5.638
	$p=.445$	$p=.216$	$p=.445$
Ambisonic $_{\text{Conceptual}}$	$z=8.500$	$z=14.333$	$z=9.333$
	S.E. = 4.717	S.E. = 5.043	S.E. = 5.043
	$p=.216$	$p=.045$	$p=.216$

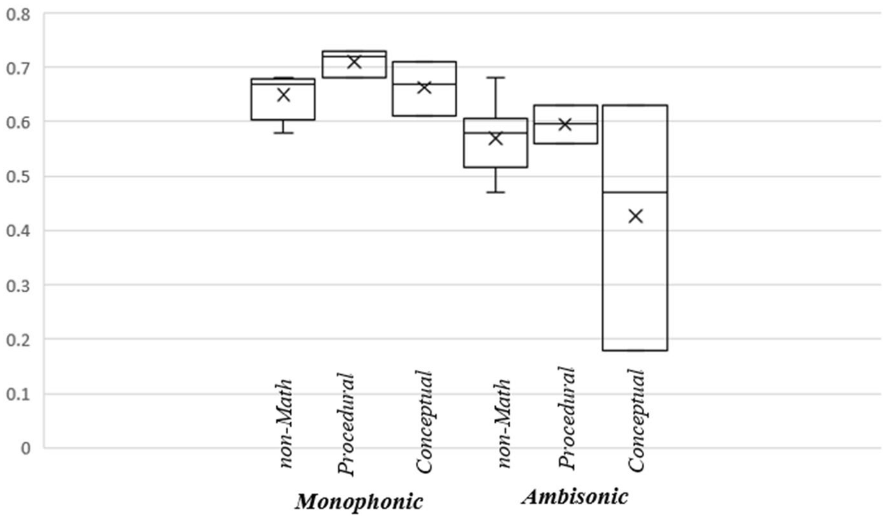

Fig. 3 Box-and-Whisker plots of participants' U_2 scores across conditions

Supplemental analysis

A benefit of convergent mixed methods designs is the ability to allow qualitative and quantitative findings to inform each other and spur deeper understanding of the data. In this study, this allowance was used to examine participants' viewing behaviors as a means to better understand the statistical results. For example, it was possible to look across individuals' change in FOV across groups in terms of what region they focused on at different points in the 360 video. Figure 4 illustrates six participants (one per comparison group). When examining across all 21 participants and considering similarities and differences within and across all six groups, two time periods stood out (2:28–2:48 and 4:56–5:20). Each of these segments were portions of whole class discussions facilitated by the recorded teacher in which students shared their approaches to finding equivalent fractions. For the segment 2:28–2:48, the teacher was in Region B and two students in Region A were describing their approach to finding an equivalent fraction. In this interaction, these two

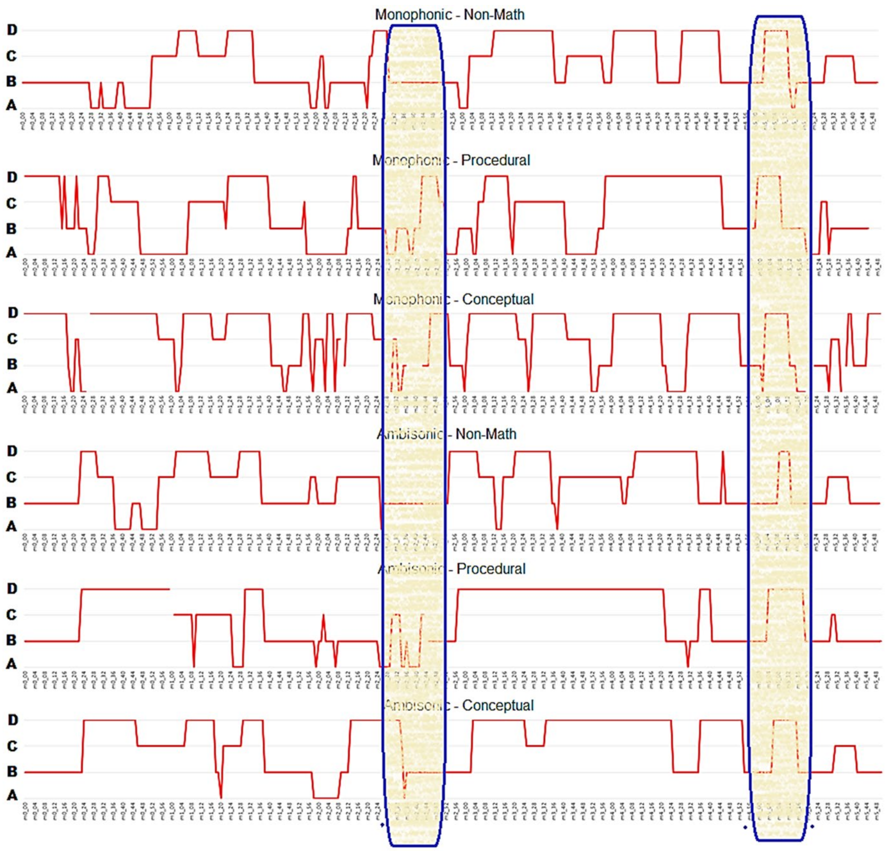

Fig. 4 Example participants viewing by section and time

students shared the rationale of not dividing because the denominator was prime. For segment 4:56–5:20, the teacher was again in Region B but was interacting with a student in Region D describing her approach to finding an equivalent fraction. Here, the recorded student described using multiplication to find a common denominator.

After identifying segments using individuals' viewing patterns, such as those illustrated in Fig. 4, individual participants' recorded viewing session were reviewed in correspondence with their coded noticing behavior (not math specific, procedural, conceptual) and variance in their FOV.

In segment 2:28–2:48, PSTs who attended to the mathematics tended to shift their focus between the teacher in Region B and the students describing their mathematics in Region A. Notably, PSTs who did not attend to mathematics (ambisonic or monophonic) focused predominately on the teacher. Participating PSTs who attended to students' procedures tended to shift their focus between the teacher and students at different tables (Regions A, C, & D) with monophonic participants demonstrating more rapid and frequent shifts in FOV. Participants attending to students' rationales and reasoning (conceptual mathematics) generally were observed to focus on students in Region D during this time, but interestingly shifted their focus to students in Region A for a few seconds. In Fig. 4, this is illustrated by the ambisonic participant who focused on Region A briefly as the two students referenced the prime number before turning to the teacher's facilitation of the discussion.

A slightly different pattern emerged for the second segment (5:04–5:12) when, following the teacher's prompting, a student in Region D shared her group's equivalent fraction of 6/16. Participants who attended to the mathematics (procedurally or conceptually) focused exclusively on this student. Interestingly, those PSTs who had monophonic audio shifted to look at the teacher writing on the board much sooner than their ambisonic counterparts. Although participants focused on more generic factors (non-math attenders) were observed to focus on Region D for a period of time, it tended to be for a shorter duration and tended to shift across multiple students at the table in Region D.

To summarize, participants in the ambisonic condition tended to shift their focus less often than participants in the monophonic condition. However, this appeared to interact with whether and how PSTs attended to the mathematics of the lesson in their written reflections. Specifically, participants who attended to students' procedures in the monophonic condition had a larger unalikeability statistic (i.e., nonparametric variance) than any other group of participants. Further, the pattern observed in Fig. 3 corresponded with supplemental analysis of PSTs' recorded FOV. Specifically, participants who attended to children's mathematical procedures tended to shift their perspective more frequently from one group of students to another, but monophonic participants did so at a much higher rate. Participants focusing on children's rationales and mathematical concepts had lower variance in their changing FOV, with ambisonic participants having the lowest variance in this regard.

Discussion

This research study set out to ask and answer two important questions about the use of ambisonic audio with immersive XR video for teaching and learning. First, the study was created to determine if ambisonic audio had any impact on the noticings or variability of focus of PSTs as they watched a 360 immersive video of a math lesson. Results provided evidence that there was a statistically significant difference between ambisonic and monophonic audio conditions. More specifically, those in the ambisonic audio conditions had significantly greater focus within the watching environment. While that finding is promising, the data did not show that the audio condition was related to whether a PST noticed important math strategies in the classroom. In other words, the use of ambisonic audio improved a users' variability of focus but did not guarantee they would always be focused on the right things within the immersive video.

Improved focus is critically important for many fields including teacher education; it is correlated to higher immersion and experience (Gandolfi et al., 2021b). For instance, Cortina et al. (2015) found that teachers who distributed their attention (as measured via eye-tracking) less evenly across students had higher productivity ratings and also provided "more elaborate and useful feedback to students" (p. 399). And while this focus does not immediately translate into improved noticing (Kosko et al., 2021b), using the technology to support focus could then lead into further instructional strategies to partner focus *with* noticing (see Kosko et al., 2022). More research is needed in this area, particularly given the findings from this study on the affordances of ambisonic audio in XR environments as they relate to focus.

The second research question asked whether there was any interaction between professional knowledge and audio condition (ambisonic or monophonic) while PSTs watched the immersive math teaching video. Professional knowledge for this inquiry was measured by whether or not they attended to the math strategies that *students* engaged with in the recorded XR video, as more knowledgeable teachers focus on students' mathematics rather than the teacher (Jacobs et al., 2010; van Es et al., 2017). For those who did not notice or attend to important math components, the audio condition did not matter. In other words, there was no statistically significant difference between ambisonic and monophonic groups for those who did not notice math strategies. This finding may be related to the fact that PSTs whose knowledge is not aligned with what was watched tend to have a lower focus in XR environments (Gandolfi et al., 2021a). Therefore, deploying immersive audio may not be sufficient to counterbalance this trend.

However, there was a statistically significant difference between ambisonic and monophonic groups in those who attended to mathematics in the XR video. More specifically, there was more variance of focus in the monophonic group. Viewers in the monophonic group spent more time looking around, attempting to find the sources of the audio. There are two reasons this is problematic. First, hypothetically, if they had to look around less, they would have been able to focus their attention, leading to better learning or instruction (e.g., Cortina et al., 2015). This ability would mimic what research says about the viewing habits of expert teachers, who tend to show a more coherent and focused viewpoint (e.g., Huang et al., 2021; Lee & Tan, 2020). In addition, this finding echoes what the literature says about ambisonic audio and its importance in making mediated environments easier to navigate and supervise for users (e.g., Chen et al., 2020; Gandolfi & Clements, 2019).

Second, and perhaps more importantly, ambisonic audio would promote less physical wear on the body during immersive video watching (and potentially during later live viewing sessions based on this practice). This relates to the proverbial concept of high-quality teachers having *eyes in the back of their head*; they do not have to look around as much to be a quality instructor. Interestingly, although not statistically significant given the current sample size, participants who attended to students' mathematics in the ambisonic condition had the lowest variance in their FOV focus of the sample. These findings need to be confirmed by further investigations, but they may suggest that high-quality teachers are better described as using their ears to see behind them.

The reduction of physical fatigue is important because it increases the amount of time that someone can spend within the virtual environment (Iskander et al., 2018). More time in the environment could lead to greater content growth and understanding. Fatigue is also one of the leading problems reported by new teachers (Bezzina et al., 2004). The purpose of this study was not to tie noticing or focus to physical and social fatigue of teachers (see Carey-Webb, 2001; Roulston et al., 2005). However, in addition to being an area for future research, there is a research-based connection between fatigue and cognitive overload (Souchet et al., 2022). It could be hypothesized that increased physical activity watching XR videos (i.e., due to the monophonic condition) could lead to cognitive overload and a lack of growth in the chosen content area.

Limitations

This article presents four main limitations. First, it addresses a specific content area (elementary math education) and, therefore, additional studies are required to problematize our findings in different domains, even beyond PST education itself (e.g., professional development, K-12 education). Second, the sample size was limited in terms of numbers and further data collections are needed to expand our conclusions, also including different populations and institutions. Third, ambisonic audio's role was evaluated in 360 videos experienced through a flat screen (e.g., participants' laptop or desktop computer); however, other immersive technologies like AR and VR may empower the potential of this feature. Finally, participants used headphones to listen to ambisonic audio. Future research could look at other ambisonic delivery techniques, modes, and settings to determine if and how they impact ambisonic audio reception.

Conclusions and future research

This research study provided two key findings that impact both theory and research. First, ambisonic audio with immersive video benefitted PSTs with professional knowledge to attend to content at hand because they don't have to look around as much. Less immersive environments could tax expertise more since those individuals may attempt to compromise between what is perceivable in the VR context and what is perceivable in real life. Thus, one facet of immersive VR, at least in the context of professional education of teachers, appears to be improving how well a representation approximates the human senses that indicate a sense of being there. The present study examined this in the context of audio (ambisonic vs monophonic).

Second, ambisonic audio decreased the overall variance in FOV focus when compared to monophonic audio viewing, regardless of the users' demonstrated professional knowledge (i.e., what they noticed pedagogically). While the ambisonic audio condition did not increase the likelihood PSTs would focus on students' math, decreasing the variance in where one looks may lend itself to supporting meaningful scaffolds to increase focusing on the content at hand (i.e., students' mathematics). Theoretically, it may also help to focus the attention of PSTs since monophonic participants who did not attend to math were looking at random places or events in the immersive video, including an observer in the back of the room, objects on the classroom wall/ceiling, and so forth.

Both overarching findings are important considering that one of the main advantages of ambisonic audio is the accessibility it provides. In other words, it provides access for those with multiple hearing needs and preferences and, therefore, serves different types of users and profiles (Rumsey, 2020). As such, the results from this study are useful pedagogically (e.g., improving PST education) and technologically (e.g., ways to reach all users).

Although results presented here are both promising and informative, future research is needed to better understand the role of audio, and other physiological factors, in how individuals engage in immersive media. One promising line of research is to examine the role of fatigue in XR environments, particularly focusing on physical and mental exertion, cognition overload, and professional noticing. Moreover, there is a need to better understand the role other senses play as interacting audio. For example, movement about a context (such as walking around a classroom) may be mimicked through multiple 360 cameras recording the same event for a multi-perspective 360 video (Zolfaghari et al., 2020). Examining the how audio (ambisonic vs monophonic) and camera placement interact may allow for a better understanding of each in XR.

Additionally, and as reported above, our study relied on using a flat screen for viewing 360 videos due to COVID-19 related limitations; further inquiries should explore the relationship between use of more immersive devices such as VR headsets and ambisonic audio, which may point at further insights and possibilities not covered in this article. The present study contributes to the small but growing body of literature focusing on the role of audio in professional education, but more work is needed to better understand and make use of audio in effective ways.

Acknowledgements Research reported here received support from the National Science Foundation through DRK-12 Grant #1908159. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Data Availability The participants of this study did not give written consent for their data to be shared publicly, so the supporting data is not available.

References

Association of Mathematics Teacher Educators [AMTE]. (2017). *Standards for preparing teachers of mathematics.* Retrieved from amte.net/standards.

Barnhart, T., & van Es, E. (2015). Studying teacher noticing: Examining the relationship among pre-service science teachers' ability to attend, analyze and respond to student thinking. *Teaching and Teacher Education, 45*, 83–93. https://doi.org/10.1016/j.tate.2014.09.005

Bezzina, C., Bezzina, N. R., & Stanyer, R. (2004). Exploring beginning teachers' perceptions of their preparation and professional development in Malta. *Mediterranean Journal of Educational Studies, 9*(2), 39–70.

Brandmeyer, A., Swedlow, N., Hertensteiner, M., & Crum, P. (2021). Benefits of immersive spatial audio on athlete reaction times in e-sports. In *Yokohama '21* (pp.1-4), May 08–13, 2021, ACM. https://doi.org/10.1145/1122445.1122456

Brunvand, S., & Fishman, B. (2007). Investigating the impact of the availability of scaffolds on preservice teacher noticing and learning from video. *Journal of Educational Technology Systems, 35*(2), 151–174. https://doi.org/10.2190/L353-X356-72W7-42L9

Buchbinder, O., Brisard, S., Butler, R., & McCrone, S. (2021). Preservice secondary mathematics teachers' reflective noticing from 360-degree video recordings of their own teaching. *Journal of Technology and Teacher Education, 29*(3), 279–308.

Carey-Webb, A. (2001). *Literature and lives*. National Council of Teachers of English.

Chen, C., Jain, U., Schissler, C., Gari, S. V. A., Al-Halah, Z., Ithapu, V. K., Robinson, P., & Grauman, K. (2020). Soundspaces: Audio-visual navigation in 3d environments. In *European Conference on Computer Vision* (pp. 17–36). Springer.

Chen, C. H. (2020). AR videos as scaffolding to foster students' learning achievements and motivation in EFL learning. *British Journal of Educational Technology, 51*(3), 657–672. https://doi.org/10.1111/bjet.12902

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education, 13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Cranley, N., Perry, P., & Murphy, L. (2006). User perception of adapting video quality. *International Journal of Human-Computer Studies, 64*(8), 637–647. https://doi.org/10.1016/j.ijhcs.2005.12.002

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Dessus, P., Cosnefroy, O., & Luengo, V. (2016). Keep your eyes on 'em all!: A mobile eye-tracking analysis of teachers' sensitivity to students. In *European conference on technology enhanced learning* (pp. 72–84). Springer. https://doi.org/10.1007/978-3-319-45153-4_6
Dong, P., & Li, A. (2021). Simulation of physical education teaching based on FPGA and wearable VR equipment. *Microprocessors and Microsystems, 81*, 103773.
Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*(3), 241–252. https://doi.org/10.2307/1266041
Fela, R. F., & Zacharov, N. (2020, May). Towards a perceived audiovisual quality model for immersive content. In *2020 twelfth international conference on quality of multimedia experience (QoMEX)* (pp. 1–6). IEEE.
Ferdig, R. E., & Kosko, K. W. (2020). Implementing 360 video to increase immersion, perceptual capacity, and teacher noticing. *TechTrends, 64*, 849–859. https://doi.org/10.1007/s11528-020-00522-3
Ferdig, R. E., Kosko, K. W., & Gandolfi, E. (2020). Effect and influence of ambisonic audio in viewing 360 video. *Journal of Virtual Worlds Research, 13*(2–3), 1–14.
Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93–106.
Gandolfi, E., Austin, C., Heisler, J., & Zolfaghari, M. (2021a). Immersive presence for future educators: Deconstructing the concept of presence in extended reality environments for preservice teachers. *Journal of Technology and Teacher Education, 29*(3), 339–367.
Gandolfi, E., Kosko, K. W., & Ferdig, R. E. (2021b). Situating presence within extended reality for teacher training: Validation of the eXtended Reality Presence Scale (XRPS) in preservice teacher use of immersive 360 video. *British Journal of Educational Technology, 52*(2), 824–841. https://doi.org/10.1111/bjet.13058
Gandolfi, E., & Clements, R. (2019). Alternative embodied cognitions at play: Evaluation of audio-based navigation in virtual settings via interactive sounds. *The Journal of Virtual Worlds Research, 12*(1), 1–14.
Gandolfi, E., Ferdig, R. E., & Kosko, K. W. (under review). Preservice teacher presence in eXtended Reality: Exploring the role of delivery platform and content area in immersive learning environments. *British Journal of Educational Technology.*
Gold, B., & Windscheid, J. (2020). Observing 360-degree classroom videos–effects of video type on presence, emotions, workload, classroom observations, and ratings of teaching quality. *Computers & Education*. https://doi.org/10.1016/j.compedu.2020.103960
Grub, A. S., Biermann, A., & Brünken, R. (2020). Process-based measurement of professional vision of (prospective) teachers in the field of classroom management: A systematic review. *Journal for Educational Research Online, 12*(3), 75–102.
Gupta, R., Lam, B., Hong, J., Ong, Z. T., Gan, W. S., Chong, S. H., & Feng, J. (2017, July). 3D audio AR/VR capture and reproduction setup for auralization of soundscapes. In *Proceedings of the 24th Intl. Congress on Sound and Vibration*, ICSV24.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Hodder Education.
Hargett, J. L. (2018). Podcasting in nursing education: Using commercially prepared podcasts to spark learning. *Teaching and Learning in Nursing, 13*(1), 55–57.
Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2021). Teachers' professional vision in action. *Zeitschrift Für Pädagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000313
Iskander, J., Hossny, M., & Nahavandi, S. (2018). A review on ocular biomechanic models for assessing visual fatigue in virtual reality. *IEEE Access, 6*, 19345–19361.
Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202.
Jong, M. S. Y., Tsai, C. C., Xie, H., & Kwan-Kit Wong, F. (2020). Integrating interactive learner-immersed video-based virtual reality into learning and teaching of physical geography. *British Journal of Educational Technology, 51*(6), 2064–2079.
Kader, G. D., & Perry, M. (2017). Variability for categorical variables. *Journal of Statistics Education*. https://doi.org/10.1080/10691898.2007.11889465
Klepper, R. K. (2005). *Silent films, 1877–1996: A critical guide to 646 movies*. McFarland.
Kosko, K. W., Ferdig, R. E., & Roche, L. (2021a). Editorial: Conceptualizing a shared definition and future directions for extended reality (XR) in teacher education. *Journal of Technology and Teacher Education, 29*(3), 257–278.
Kosko, K. W., Ferdig, R. E., & Zolfaghari, M. (2021b). Preservice teachers' professional noticing when viewing standard and 360 video. *Journal of Teacher Education, 72*(3), 284–297. https://doi.org/10.1177/0022487120939544

Kosko, K. W., Heisler, J., & Gandolfi, E. (2022). Using 360-degree video to explore teachers' professional noticing. *Computers and Education, 180*, 1–13. https://doi.org/10.1016/j.compedu.2022.104443

Landis, J. R., & Kotch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, A. V. Y., & Tan, S. C. (2020). From micro to meso: Scaling of a teacher noticing study. In H. J. So, M. M. Rodrigo, J. Mason, & A. Mitrovic (Eds.), *Proceedings of the 28th international conference on computers in education* (pp. 100–105). Asia-Pacific Society for Computers in Edication.

Li, D., Langlois, T. R., & Zheng, C. (2018). Scene-aware audio for 360 videos. *ACM Transactions on Graphics (TOG), 37*(4), 1–12.

Malham, D. G., & Myatt, A. (1995). 3-D sound spatialization using ambisonic techniques. *Computer Music Journal, 19*(4), 58–70.

Masataka, N., & Perlovsky, L. (2012). The efficacy of musical emotions provoked by Mozart's music for the reconciliation of cognitive dissonance. *Scientific Reports, 2*(1), 1–4.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389–401. https://doi.org/10.1037/0022-0663.86.3.389

Mota, P., Carvalho, N., Carvalho-Dias, E., Costa, M. J., Correia-Pinto, J., & Lima, E. (2018). Video-based surgical learning: Improving trainee education and preparation for surgery. *Journal of Surgical Education, 75*(3), 828–835.

Narbutt, M., Allen, A., Skoglund, J., Chinen, M., & Hines, A. (2018, May). AMBIQUAL-a full reference objective quality metric for ambisonic spatial audio. In *2018 tenth international conference on quality of multimedia experience (QoMEX)* (pp. 1–6). IEEE.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Poquet, O., Lim, L., Mirriahi, N., & Dawson, S. (2018, March). Video and learning: a systematic review (2007–2017). In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 151–160).

Prakash, S. S., Muthuraman, N., & Anand, R. (2017). Short-duration podcasts as a supplementary learning tool: Perceptions of medical students and impact on assessment performance. *BMC Medical Education, 17*(1), 1–14.

Reddy, A. P., & Vijayarajan, V. (2020). Audio compression with multi-algorithm fusion and its impact in speech emotion recognition. *International Journal of Speech Technology, 23*(2), 277–285.

Roulston, K., Legette, R., & Womack, S. T. (2005). Beginning music teachers' perceptions of the transition from university to teaching in schools. *Music Education Research, 7*(1), 59–82.

Rumsey, F. (2020). Immersive audio—Defining and evaluating the experience. *Journal of the Audio Engineering Society, 68*(5), 388–392.

Schack, E. O., Fisher, M. H., Thomas, J. N., Eisenhardt, S., Tassell, J., & Yoder, M. (2013). Prospective elementary school teachers' professional noticing of children's early numeracy. *Journal of Mathematics Teacher Education, 16*(5), 379–397.

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Scwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education, 27*(2), 259–267.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). McGraw-Hill Inc.

Sihotang, R. A., Cendana, W., & Kristidhika, D. C. (2021). The use of video in improving students' attention in learning process of kindergarten students. *International Journal of Elementary Education, 4*(4), 496–502. https://doi.org/10.23887/ijee.v4i4.28430

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika, 73*(3), 751–754. https://doi.org/10.2307/2336545

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human-Computer Interaction, 38*(9), 801–824.

Teuscher, D., Leatham, K. R., & Peterson, B. E. (2017). From a framework to a lens: Learning to notice student mathematical thinking. In E. O. Schack, M. H. Fischer, & J. A. Wilhelm (Eds.), *Teacher noticing: Bridging and broadening perspectives, contexts, and frameworks* (pp. 31–48). Springer.

Ulrich, F., Helms, N. H., Frandsen, U. P., & Rafn, A. V. (2021). Learning effectiveness of 360 video: Experiences from a controlled experiment in healthcare education. *Interactive Learning Environments, 29*(1), 98–111.

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.
Walshe, N., & Driver, P. (2019). Developing reflective trainee teacher practice with 360-degree video. *Teaching and Teacher Education, 78*, 97–105. https://doi.org/10.1016/j.tate.2018.11.009
Weston, T. L., & Amador, J. M. (2021). Investigating student teachers' noticing using 360 video of their own teaching. *Journal of Technology and Teacher Education, 29*(3), 309–338.
Yang, K., Russell, B., & Salamon, J. (2020). Telling left from right: Learning spatial correspondence of sight and sound. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9932–9941).
Yao, S. N. (2017). Headphone-based immersive audio for virtual reality headsets. *IEEE Transactions on Consumer Electronics, 63*(3), 300–308.
Yip, J., Wong, S. H., Yick, K. L., Chan, K., & Wong, K. H. (2019). Improving quality of teaching and learning in classes by using augmented reality video. *Computers & Education, 128*, 88–101.
Zolfaghari, M., Austin, C. K., Kosko, K., & Ferdig, R. E. (2020). Creating asynchronous virtual field experiences with 360 video. *Journal of Technology and Teacher Education, 28*(2), 315–320.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Richard E. Ferdig is the Summit Professor of Learning Technologies at Kent State University. His research focuses on the development of innovative technologies for teaching and learning, online and blended learning, the relationship between faith and technology.

Karl W. Kosko is an associate professor in mathematics education at Kent State University. His research focuses on how mathematical meaning is conveyed through media, representations, and linguistic discourse.

Enrico Gandolfi is an assistant professor in Educational Technology at Kent State University. His research interests include online communities of inquiry on social media and the learning potential of extended reality technologies.