Coordinating Modalities of Mathematical Collaboration in Shared VR Environments

Wen Huang, Southern Methodist University, wenhuang@smu.edu
Candace Walkington, Southern Methodist University, cwalkington@smu.edu
Mitchell J. Nathan, University of Wisconsin-Madison. mnathan@wisc.edu

Abstract: This study investigates how learners collaboratively construct embodied geometry knowledge in shared VR environments. Three groups of in-service teachers collaboratively explored six geometric conjectures with various virtual objects (geometric shapes) under the guidance of a facilitator. Although all the teachers were in different physical locations, they logged into a single virtual classroom with their respective groups and were able to see and manipulate the same geometric shapes as well as see their collaborators' avatars and actions on the shapes in real time in the shared virtual space. This paper introduces a novel multimodal data analysis method for analyzing participants' interactive patterns in collaborative forms of actions, gestures, movements, and speech. Results show that collaborative speech has a strong simultaneous relationship with actions on virtual objects and virtual hand gestures. They also showed that body movements and positions, which often focus on virtual objects and shifts in these movements away from or around the object, often signal key interactional collaborative events. In addition, this paper presents five emergent multimodality interaction themes showing participants' collaborative patterns in different problem-solving stages and their different strategies in collaborative problem-solving. The results show that virtual objects can be effective media for collaborative knowledge building in shared VR environments, and that structured activity design and moderate realism may benefit shared VR learning environments in terms of equity, adaptability, and cost-effectiveness. We show how multimodal data analysis can be multi-dimensional, visualized, and conducted at both micro and macro levels.

Keywords: multimodality; mathematics education; virtual reality; collaborative learning; embodied learning

Introduction

Theories of embodied learning emphasize how all human reasoning is embedded within perception, spatial systems, social settings, and physical motions like gestures. Embodied learning research in mathematics has increasingly made use of novel technologies promoting important physical actions that embody mathematical ideas, such as motion capture or video games (Abrahamson et al., 2020). Less research has investigated how math learning can be embodied in VR environments, perhaps because "bodies" in this context are virtually rendered and may be seen as inherently "disembodied." VR environments that afford meaningful embodied collaboration where learners are able to use their bodies together in concert to explore ideas have also been relatively rare. Here we investigate a collaborative VR environment for geometric reasoning and explore novel ways in which learners use their virtual bodies to physically instantiate, highlight, and explore mathematical ideas together.

Virtual reality (VR) is a fully immersive 3D multimedia environment that people can interact with in a realistic manner that mimics real or imagined experiences. VR may be especially beneficial for presenting 3D objects in a manner that supports collaborative embodied interactions, as learners can interact with 3D objects in a 3D world. New forms of VR enable multiple learners to manipulate *the same* mathematical shapes presented as *dynamic objects* projected in a joint 3D collaborative space, using intuitive *hand gestures*. Johnson-Glenberg (2018) gives three dimensions of effective embodied learning using VR interventions: motoric engagement, gestural congruency, and perception of immersion. We built upon these ideas to construct our collaborative VR environment.

It is through gestures, body movements, perspective-taking, talk moves, and actions on objects that mathematics learning becomes embodied in collaborative settings. For example, according to Walkington et al. (2019), gestures can be used for embodied collaboration —students echo and mirror each other's gestures, make new gestures to differentiate an idea that appeared in another's gestures, and conjointly gesture to form mathematical objects. During collaboration, students gesture in *fundamentally* different ways than when alone, which reveals the *distributed* nature of their co-constructed cognitive processes. Furthermore, discussions in computer-supported collaborative learning (CSCL) environments also involve particular kinds of talk moves used to represent mathematical ideas, establish shared understanding, engage in negotiation, and plan future actions (Andrews-Todd et al., 2019). Beyond talk moves and gestures, embodied collaboration also emerges through learners' manipulation of shared virtual objects, and the ways in which learners position their virtual avatars' bodies; these have been highlighted less often in math studies. Thus, developing a scheme for better understanding these embodied moves is a goal of the present study.

This study aims to understand how learners collaboratively construct embodied knowledge in VR environments. We observed in-service teachers who were enrolled in a virtual course that required them to grapple with 3D geometry concepts. It is expected that the identified collaborative patterns can partially direct the design and use of shared VR environments in mathematics education.

Theoretical framework

The theoretical framing of this study integrates theories of embodied cognition with theories of distributed and extended cognition. We focus on four different modalities for embodied collaboration in shared VR environments (i.e., action on virtual objects, gesture, body movement, and speech), in the context of geometric reasoning. Following is a review of these theories, concepts, and ideas.

Embodied cognition

Recent research in the learning sciences suggests that experiences that sustain learning are enacted, embedded, extended, and embodied, or "4E" (Newen et al., 2018). Theories of embodied cognition conjecture that human cognitive processes are deeply rooted in the body's interactions with the world (Wilson, 2002). This is at odds with a view of mathematics as a discipline disconnected from the body and from action, based on abstract formalisms that have few real-world referents (Lakoff & Núñez, 2000; Nathan, 2012). Johnson-Glenberg and Megowan-Romanowicz (2017) proposed a taxonomy of embodiment in education. According to the taxonomy, the degree of embodiment of an educational intervention is predicated by a) sensorimotor engagement, b) the congruency between the gestures and the content to be learned, and c) the amount of immersion. They suggest that higher levels of embodiment are associated with higher levels of learning from these environments.

Theories of embodied cognition have inspired researchers to explore methods of using bodies/actions for mathematics teaching and learning (Abrahamson et al., 2020; Georgiou & Ioannou, 2019). Previous research has provided evidence to support the affordances of body-based tasks in helping students/kids develop mathematical understanding about angles (Smith et al., 2014), conduct algebraic derivations (Weitnauer et al., 2017), enhance mathematical imagination around ideas of ratio and proportion (Abrahamson & Sánchez-García, 2016), construct justifications for geometry conjectures (Walkington et al., 2022), and estimate the mental number line (Fischer et al., 2014), among many others.

Distributed and extended cognition

Hutchins (1995, 2000) challenged the traditional view of using the individual person as the unit when analyzing cognition. He proposed that the central unit of analysis is the functional system, which includes individuals, artifacts, and their relationship in a particular context. Cognitive processes can be distributed across collaborators and between individuals and external environments. Cognitive activities take place via the propagation of the representational state across internal (i.e., individual memories) and external (e.g., computer and paper) representations through various communicative processes (e.g., speech, operations, and the construction of artifacts) (Rogers & Ellis, 1994). Media and representations (e.g., computer and paper-based displays) support knowledge sharing in coordinating interdependent activities (Rogers & Ellis, 1994).

Clark and Chalmers (1998) similarly believed the surrounding features of an environment can play a crucial role in driving an individual's cognitive processes. This enables an individual's cognition to be expanded beyond the brain into the environment. Sutton and colleagues (2010) claimed that extended cognition does not only mean a parity between or functional isomorphism

of neural and extra-neural features; it is the complementarity between our inner and external heterogeneous resources that creates the extended cognitive system. In mathematics, Walkington et al. (2019) investigated the potential of learners using their bodies to discuss and explore mathematical conjectures with a shared goal. They found that learners used collaborative gestures to explore these conjectures that stretched over multiple peoples' bodies. They claimed that collaborative gestures can facilitate and extend cognitive processing in a distributed cognitive system.

CSCL as a field has been traditionally focused on group *intersubjectivity* – or the establishment of a shared understanding and advancement of knowledge among learners in CSCL environments (Stahl, 2015). However, research from the perspective of embodied cognition typically focuses on the behaviors of individuals. In our view, learning in CSCL cannot be fully understood as the sum of individual thoughts or actions, but rather occurs at the group level, through the knowledge captured by group products and artifacts (which may not be reflective of any individual's knowledge). Meaning, then, arises as actions performed by individuals are combined and considered in the context of the entire situated group as the interactional unit. This focus on group-level phenomena is well-suited to theories of distributed and extended cognition that offer broader notions of cognition, knowledge, and goal-directed action. Despite this approach being foundational to CSCL as a field, this kind of group-level analysis has been difficult to consistently conceptualize and enact (Stahl, 2015), making the present investigation well-suited to the field.

Multimodal interaction

Multimodality is a term that has been used in many different ways (Jewitt et al., 2016). Within perspectives inspired by gesture studies and the study of social interaction, such as conversation analysis (CA), multimodality refers to the various (semiotic) resources mobilized by participants for organizing their actions, such as gesture, gaze, body posture, and movement (Mondada, 2014, 2016). Each resource offers distinct possibilities and limitations (Jewitt et al., 2016). However, when joined together in local contexts of action, these diverse resources have the potential to create a whole that is both greater and different from any of its constituent parts (Goodwin, 2000). Simultaneity and sequentiality (i.e., interactional order, whether actions occur at the same time or one after another) are the fundamental principles of multimodal interaction (Mondada, 2016). The CA framing of multimodality highlights the notion that people build action with different semiotic resources (Jewitt et al., 2016). These resources are constitutively intertwined (Mondada, 2014), and the relationship between the simultaneity of different modalities and the sequentiality of activities is complex. Asynchronicities between modalities do not seem to be accidental. Activities performed in one modality may not have the exact boundaries of action as those performed in another modality (Deppermann, 2013). Additionally, sequentiality may not be organized strictly successively; it relies on the prior and subsequent actions in real time, and coordinated, simultaneous multimodal interactions are intertwined (Mondada, 2016). Furthermore, actions are organized not only by individual speakers but also within social interaction (Mondada, 2016). Thus, we must consider co-participants' simultaneous multimodal activities and managing action sequences in CA (Deppermann, 2013; Schegloff, 2007). We use these principles from multimodality to frame the current investigation.

Geometric Reasoning

Geometry emerged as people have strived to measure (-*metry*) and know the world (*geo*-). It is now recognized as central to intellectual inquiry and design of nearly every facet of human life, including politics, art, games, biology, and machine learning (Ellenberg, 2021). School geometry courses study spatial objects, relationships, transformations, and the corresponding mathematical systems that represent them (Clements & Battista, 1992). A common experience in these courses is where students name parts of 2D and 3D objects, transform objects through rotation/perspective and scaling, object construction, and express justifications and proofs to convince themselves and persuade others a conjecture is true or false (Harel & Sowder, 2007). The *proof* is a mathematical argument including a connected series of assertions that are valid and commonly accepted by the community (Stylianides, 2007). Traditionally, two-column proofs where students write "Statements" and "Reasons" have dominated, but there is an increasing recognition that students need to engage with proofs in more diverse ways to fully engage with mathematical ideas (Cirillo & Herbst, 2012; Herbst, 2002).

Justifications support mathematical claims by explaining why claims make sense, providing insight into the underlying mathematics (Bieda & Staples, 2020; Staples et al., 2017). Justifications may not be logically complete or mathematically exhaustive and thus may not be universally acceptable to a classroom community in the ways proofs are. However, justifications can be an important step toward proofs (Staples et al., 2017) and can help students develop mathematical ideas (Ellis et al., 2022). Engaging in justification and proof practices can benefit students' mathematical understanding (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices, 2010), and these practices can be expressed in various formats and modalities. Students may present justifications and proofs through spoken or written language (Harel & Sowder, 2007; Healy & Hoyles, 2000). Additionally, simulated actions (as expressed through gestures) may also be used to engage in justification and proof through embodied mathematical reasoning (Walkington et al., 2022), and students can build their understanding of proofs of geometric conjectures by engaging in collaborative forms of gesturing (Walkington et al., 2019).

Shared VR environments

Shared VR environments enable people to interact with collaborators in a multimodal way within the virtual environment. This is consistent with the triad structure of collaboration in CSCL where computational artifacts mediate the actions of participants (e.g., participant – artifact – participant; Ludvigsen & Steier, 2019). Many current VR simulations are synchronous and support both remote collaboration, where users are in different physical locations, and co-located collaboration. While co-located collaboration has the potential to support offline and latency-free collaboration, remote VR has the advantage of avoiding users' collisions in the physical world (Pidel & Ackermann, 2020). In addition, compared with traditional web-based online discussion, shared VR environments show several advantages. First, learners can deliver knowledge and disperse information through embodied action and social interactions (Zheng et al., 2018). Second, shared VR environments can situate learners in complex and meaningful contexts and engage learners with different perspectives to solve a range of practical problems, fostering collaboration skills

and innovation (Marky et al., 2019; Philippe et al., 2020; Zheng et al., 2018). Third, it can be easier to clearly communicate a concept with virtual VR objects, as there is less reliance on imagination in shared VR environments (Pidel & Ackermann, 2020).

Embodied forms of collaboration for shared VR

Understanding the different ways communication is embodied and jointly-constructed is a key element in unpacking *collaborative embodiment* in CSCL environments – here defined as the way learners co-create actions and embodied ideas that become extended (Clark, 2012) over multiple learners. It is through this multimodal account of communication that learners generate mathematical meaning. It is also through this account that learners establish intersubjectivity, joint attention, and shared knowledge; a critical element in solving complex tasks in CSCL (Ludvgsen & Steier, 2019). We examine four modalities for embodied communication that are particularly important in VR CSCL environments – collaborative actions on virtual objects, collaborative gestures, collaborative body movements, and collaborative speech.

    Collaborative actions on virtual objects. Meaningful, shared representations are necessary to successfully carry out many technical activity structures, such as ship navigation (Hutchins, 1995). One profound affordance of VR is gestural congruency – the notion that objects can be manipulated intuitively through hand gestures (Lindgren & Johnson-Glenberg, 2013). When gestural congruency is present, physical actions may be beneficial for recall (Glenburg et al., 2007; Suh & Moyer, 2007), and may create embodied resources and metaphors for learners (Abrahamson & Sánchez-García, 2016; Alibali & Nathan, 2012; Smith et al., 2014). Learners may prefer tangible user interfaces where objects are manipulated by gesture to more typical graphical user interfaces that use buttons and menus (Zuckerman & Gal-Oz, 2013). The agency that comes with being able to act upon the world in three dimensions is also a profound affordance of VR (Johnson-Glenburg, 2018).

    When learners manipulate mathematical objects (like triangles, lines, etc.) using gestures in VR environments, they may engage in various actions, including resizing, rotating, reflecting, constructing, and dilating. VR technologies allow learners to easily explore geometric objects at many different scales (Dimmel & Bock, 2019) – in VR, learners are able to make an icosahedron in the palm of their hand, or they could make one so large that they could walk inside of it. These manipulations all occur in a context where multiple learners can manipulate the same object at the same time and see each other's manipulations in real time. Thus, unlike in traditional simulation environments, these actions are transformed to have collaborative goals and implications. When using simulations on a flat screen, a single learner usually has individual control of the virtual content, creating challenges for collaboration. In VR, learners can see the virtual content and their collaborators' avatars in a 3D space; each person has their own perspective and can exert control over the simulation (Bujak et al., 2013; Johnson-Glenburg, 2018). This also creates an opportunity for unique collaborative moves involving learners engaging in actions on virtual objects.

    Collaborative gestures. *Gestures*, spontaneous or purposeful hand and arm movements that often accompany speech, are a powerful way simulated actions can give rise to physical movements (Hostetter & Alibali, 2008). During collaboration, gestures often operate synchronously with speech, acting as a mechanism to create cohesion and bind conversational elements together (Enyedy, 2005; Koschmann & LaBarron, 2002). Learners often perform *collaborative gestures* – jointly-construed physical movements that demonstrate mathematical

relationships with their bodies. Walkington et al. (2019) introduced a framework for specifying how learners can collaborate using gestures while solving math problems. Learners repeat others' gestures through echoing or mirroring gestures, build on one another's gestures through alternation gestures, and physically co-represent a single mathematical object using joint gestures. Across two exploratory studies, we found that in comparison with gestures that were not collaborative, gesturing collaboratively was associated with higher geometric proof performance (Walkington et al., 2019; Abrahamson et al., 2020).

Collaborative body movements. Body motions, defined as the way the learner moves their avatar (including the avatar's head) to engage in collaborative reasoning and problem-solving, are another key type of collaborative interaction in VR environments. For example, VR systems allow learners to change perspectives by moving their bodies, and as a result learners have direct access to 3D figures rather than 2D projections of 3D figures (Dimmel & Bock, 2019). Johnson-Glenburg (2018) refers to this as one of the two profound affordances of VR (see also Wu et al., 2013). Learners may engage in collaborative body movements where they contribute to problem-solving processes by taking different perspectives on the same object, or even taking the perspective of one of their collaborators to better understand their point of view. In addition, in VR environments, learners might position their bodies with respect to where their peers are standing or where virtual objects are and may move around the environment as they engage in different collaborative structures. Aiming the virtual head at a collaborator might be a way to demonstrate the learner's attention to them or to make a bid for their attention. Collaborating around shared objects within VR may be more effective than collaborating around a screen, as VR learners may not have to engage in as much split attention between the screen and their collaborators (Shelton & Hedley, 2004).

Collaborative speech. Embodied accounts of language offer valuable insights into how cognition appears to be based on sensorimotor processes. Syntax, far from being a purely formal entity, protected from the intrusions of meaning, context, culture, or evolutionary adaptation, is a product of one's biological, cultural, and physical ontology (Lakoff & Johnson, 1999). Behavioral and neural evidence shows cognitive processes for processing numbers and numerical operations are mediated by language-based processes (Dehaene, 1997). We have found that learners' speech patterns independently predict whether learners' justifications are mathematically sound (Pier et al., 2019). Frameworks for discourse analysis in CSCL (Weinberger & Fischer, 2006) suggest that when learners engage in collaborative discourse, they engage in argumentative knowledge construction. Andrews-Todd et al. (2019) created a rubric for collaborative talk moves in CSCL – including maintaining communication, sharing information, establishing shared understanding, and negotiating. Katić et al. (2009) further examine how mathematical representations can be used during collaboration as visual stimuli, an isomorphism between the task and the materials, a learning strategy or behavior, or a visual explanation for a completed idea.

Summary and Research Questions

In shared VR environments, learners collaborate by leveraging multimodal resources such as actions, gestures, body movements, and speech. In this way, learners' cognition becomes stretched over multiple virtual bodies and the objects and elements of the virtual environment itself. These interactions occur in a context where objects can be dynamic and learners can participate in joint interaction. Such environments have considerable affordances for learning about mathematical ideas involving dynamicity and transformations and for engaging in rich embodied discussions with other learners and facilitators. Research is needed to understand how learners leverage these embodied multimodal resources to build shared

understanding together in VR. The present study focuses on looking for key interactional patterns as learners explore 3D geometric properties in a shared VR environment. Our research question is: *How do learners collaboratively orchestrate embodied semiotic resources (e.g., gesture, action, holographic object use, and speech) to engage in geometric reasoning in shared VR environments?* This question includes two sub-research questions:

1. *How do groups of learners orchestrate modalities simultaneously or in sequence in shared VR environments?*
2. *How do groups of learners orchestrate modalities with collaborators in shared VR environments?*

Method

This section introduces the shared VR environment and its implementation with participants and our data analytic approach. The analytic approach includes analyses at both macro and micro levels, which each focus on answering one of two sub-research questions.

Participants

Recruitment was conducted in a virtual education course for math teachers at a private university in the Southern United States. The class was fully virtual due to the COVID-19 pandemic. All nine female in-service teachers in this course consented to participate in this study; there were no financial or external rewards for participation. These teachers had an average of 3.9 years of teaching experience. Table 1 shows their demographic information.

Table 1. Participants' demographic information

		n
Age	20-29	5
	30-39	2
	40-49	0
	50-59	2
Race/Ethnicity	Asian	3
(Multiple selections allowed)	White	6
	Hispanic	1
School Role	Grades 5-6 mathematics teacher	3
	Grade 7-10 mathematics teacher	5
	Technology lead for Grade 4-6	1

Materials

A VR simulation designed by our team, which we refer to as the Geometry Simulation Environment (GSE), was installed on Oculus Quest VR goggles, which were checked out from the

VR lab at the university by the teachers enrolled in the course. The GSE was a product of an initial design cycle of a design-based research program on mixed reality for geometry learning. Participants joined the VR simulation from their homes and were placed in a virtual classroom together after entering fake names and selecting the target classroom from a dashboard. Fixed-appearance avatars with a head, upper torso, and hands represented participants in the virtual classroom. The participant's name hovered near the head of the avatar, with the upper torso and head moving together according to the goggles' movements. Virtual hands would appear via hand-tracking when the participants' hands were in view of the goggles. A participant can see other avatars and hear their voices through the goggles' speaker, with their speech recorded through the goggles' microphone. Figure 1 shows three avatars in GSE, with the middle avatar pointing to a virtual object in the center with her left hand (highlighted by the red circle in this figure).
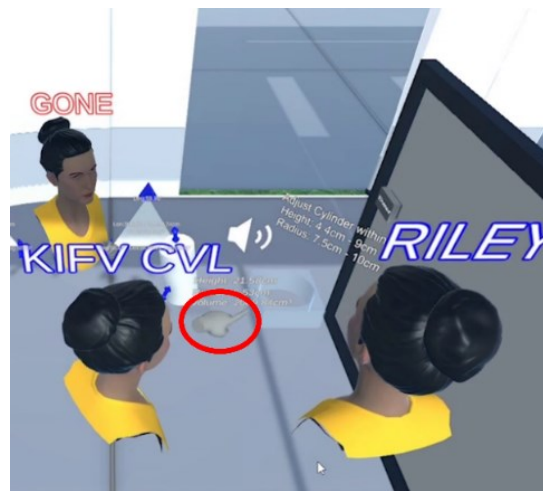

Figure 1. A discussion scenario in the GSE

Six secondary-level mathematics tasks were used to facilitate geometric discussions and explorations in the GSE. In each task, participants needed to determine the validity of a conjecture or use actions to create or highlight a mathematical representation. Table 2 shows the tasks and the corresponding mathematical concepts and skills in these tasks. In Cylinder 1 and 2 Tasks, participants could move, grasp, and modify the height and radius of a cylinder in the GSE. The critical measurements of the cylinder, such as the height, volume, and radius, were displayed and updated in real time. In Triangle 1 and 2 Tasks, participants would manipulate two virtual triangles. In this process, participants could drag the triangle's vertices to change the triangle's side lengths and angles. Similarly, the triangle's angle, side, and area were displayed and updated automatically. In Cube Task, participants would collaboratively manipulate a cube. In Volume Task, participants would resize two of the available shapes in the environment (i.e., a cube, a square pyramid, a sphere, a hexagonal prism, and a torus).

The shared objects as well as the real-time measurement data in the GSE might help participants construct justifications and proofs related to these tasks. For example, in Volume Task, a visualized comparison of the two selected solids allowed participants to adjust the size of these solids collaboratively and inspired them to use formulas to verify their judgement. Similarly, in Triangle 1 Task, participants could see the triangle's side length and the degree of each side's opposite

angle in real-time when they adjusted the triangle. Comparing the degree between the largest side's opposite angle and the other two angles in different configurations could provide evidence to support participants' reasoning. Overall, the GSE provided an embodied way for justification and proof practices to emerge. This method might be particularly beneficial when a task involves concepts related to 3D scaling, size, and volume.

Table 2. Mathematics Tasks

Task Name	Conjecture	Mathematical Concepts and Skills
Cylinder 1 Task	"One of your students conjectures that the volume of a cylinder changes by the same amount whether you increase the radius by 1 cm or increase the height by 1 cm. Do you think this conjecture is true or false? Why? Try it out with the cylinder in front of you."	3D Scaling, Volume & Justification/proof.
Cylinder 2 Task	"Can you make it so the cylinder looks like a circle from both your viewpoint and your partner(s) viewpoint(s), at the same time?"	3D perspective.
Triangle 1 Task	"One of your students conjectures that for all triangles, the largest side is always opposite from the biggest angle. Do you think this conjecture is true or false? Why? Try it out with the triangle in front of you?"	linear and angular measure & Justification/proof
Triangle 2 Task	"Can you make the two triangles into a square, with one person controlling each triangle?"	Proof by Construction.
Cube Task	"Can you and your partner(s) place your hands on as many faces of the cube as possible? Can you point your fingers to as many vertices of the cube as possible? Can you use your index finger and thumbs to cover as many edges as possible? Is it possible to cover all the edges with the number of people we have here? How many vertices, faces, and edges does a cube have?"	Properties of polyhedra
Volume Task	"Choose two of the solids. Size them such that you have created two solids that you believe would have the same volume. How do you know they have approximately the same volume?"	Volume, 3D scaling, Justification.

Procedure

Participants were assigned into three groups in the VR shared GSE. Each group engaged in a one-hour experiment with a facilitator. The role of the facilitator was to ask participants questions, solve technical issues, and control their progress. When participants first entered the virtual

classroom, they would see a cylinder and then start Cylinder Task 1. Figure 2 shows the sequence of VR learning activities. Participants had a half-hour break between the first four tasks and the remaining two. Participants were video recorded by using their own first-person VR goggles' recording feature, which is physically located in the goggles, as well as by using an additional added feature where a virtual overhead camera was recording the interactions.
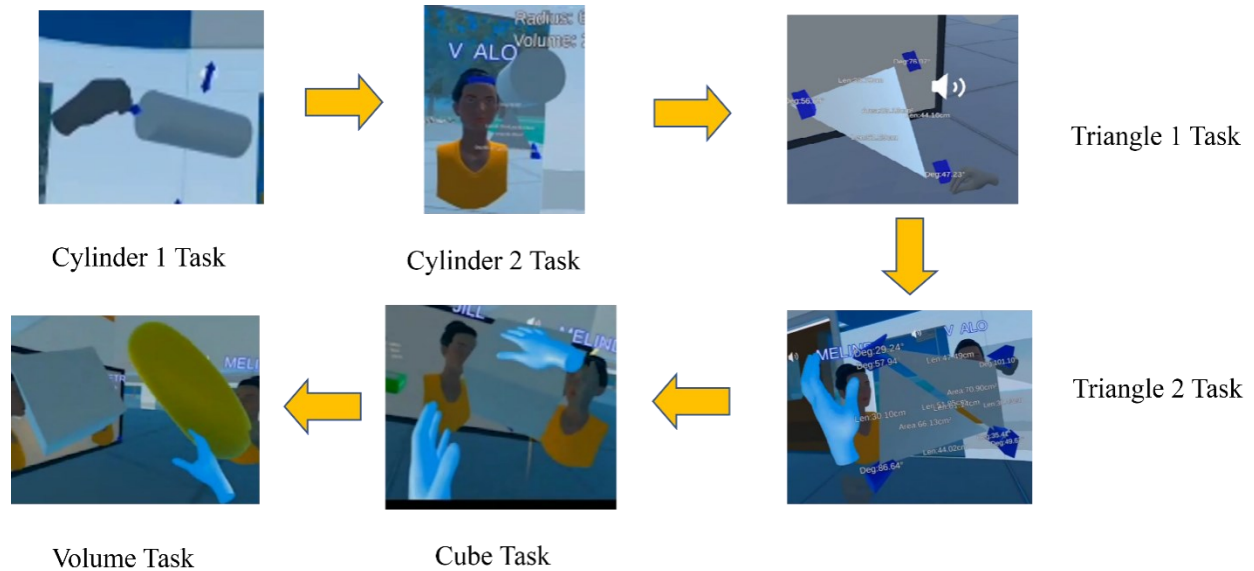

Figure 2. The sequence of VR learning activities

Data analytical approach

Researchers have explored how to conduct effective multimodal interaction analyses for decades. In earlier studies, researchers mainly focused on how to effectively present peoples' interactions on a timeline. For example, Schroeder and colleagues (2006) describe one method for analyzing interactions in collaborative virtual environments, where each person's communicative acts are quantitatively captured on a timeline. Evans and colleagues (2011) used Excel transcripts to show topic units and shared focus between participants in order to identify children's communicative strategies in solving geometric puzzles in CSCL contexts. However, researchers gradually found that using a timeline is not enough to fully support the description of interactions in collaboration. They accepted the idea of using a multi-level structure to represent the hierarchical relationship among video segments. Norris (2016) defined mediated action as the unit of analysis for video conferences and claimed that lower-level actions can be performed within the performance of higher-level actions. Hod and Twersky (2020) used a 4-level structure to divide the videos of group collaboration in an augmented reality (AR) sandbox and then analyzed how participants interacted in each video segment based on the pre-defined types of spatial actions. Overall, previous research shows that timelines and multi-level structures are effective tools in multimodal data analyses. However, missing from prior research is a practical method to guide researchers to efficiently establish the relationships between focused actions within a video segment, as well as relationships

between video segments. This gap weakens the potential of research identifying the interactive patterns in collaborative learning.

Considering previous research, our theoretical framework, and the research questions, a novel multimodal data analytical method was introduced to focus on our four key collaborative modalities (i.e., actions on virtual objects, gestures, body movements, and speech). This method allows researchers to conduct analyses at both micro and macro levels. Particularly, it simplifies the process of establishing the relationship between actions by visual modeling. As the first step, the speech from the video recordings was transcribed, and descriptions of avatars' actions, gestures, and movements were added, time-stamped, and integrated with the speech records. Then, the V-note software (Bremig LLC, 2022; see Figure 3) was used to code the video records. V-note supports direct labeling of codes on video records for specific elapsed time sequences at different grain sizes. Researchers can see the labeled video segments, text transcripts, and codes in its main interface and visually read each labeled video segment on a timeline. The first author labeled one-fourth of the video records using the open coding method (Benaquisto, 2008) and then met with the second author to finalize the coding method. Referring to the prior studies in CSCL and embodied learning, such as Andrews-Todd and colleagues' (2019) collaborative problem-solving ontology and McNeill's (1992) gesture categories, the coding results were updated and a formal codebook was developed, in which each modality has 4-5 different sub-codes. Based on this codebook, the first author started to label the whole video record, and the second author labeled one-half of the video record. The two authors had biweekly meetings to discuss the emerging labeling issues and continuously updated the codebook to reflect their consensus. Table 3 shows the final codebook. Multiple labels were allowed on a video segment. For example, a video segment showing a participant changing the height of a cylinder to answer conjecture 1 ("One of your students conjectures that the volume of a cylinder changes by the same amount whether you increase the radius by 1 cm or increase the height by 1 cm. Do you think this conjecture is true or false? Why? Try it out with the cylinder in front of you") was labeled by "resize," "measure," and "instructed action" at the same time. This is because "resize" is an "instructed action" in the task; and when resizing the cylinder, the participants measured the change of the cylinder's volume at the same time.
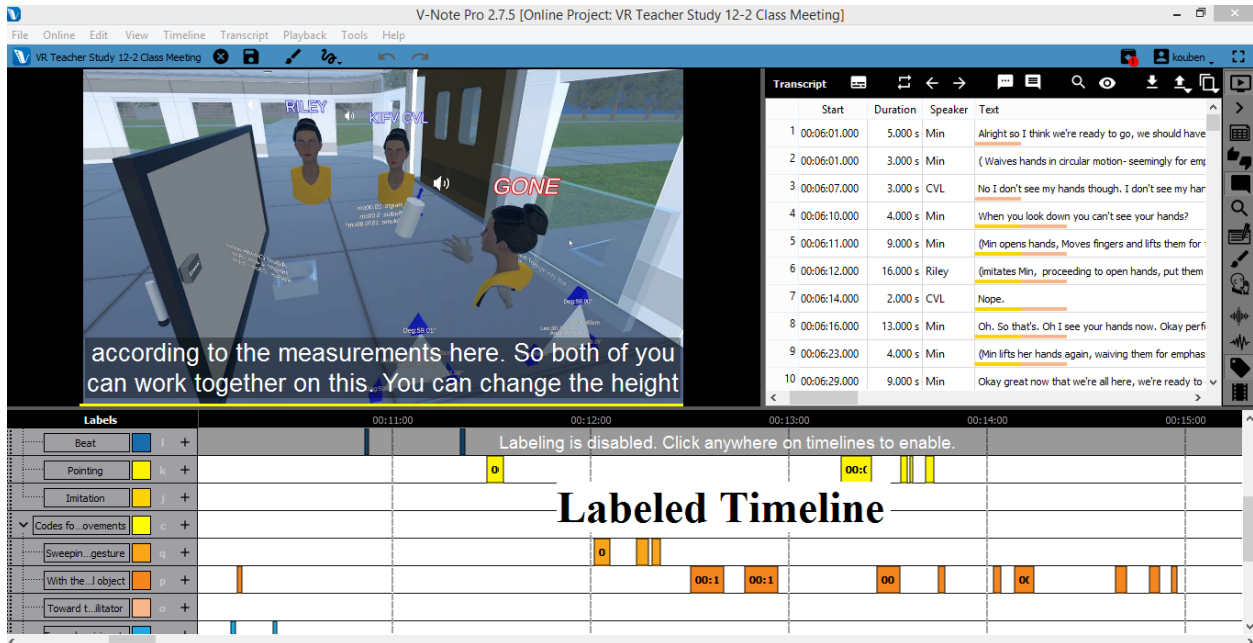
Figure 3. A screenshot of V-note software

Table 3. Codebook of the video analysis

Coding Category	Description	Explanation and Examples
1. Actions on Virtual Objects		
Instructed action	An umbrella category for actions on virtual object instructed by the facilitator or participants.	Actions instructed to complete a task, including (1) resize and measure in Cylinder 1 Task, (2) resize and move in Cylinder 2 Task, (3) resize and measure in Triangle 1 Task, (4) resize, move, and measure in Triangle 2 Task, (5) measure in Cube Task, and (6) resize and measure in Volume Task. Whether or not "miss" concurrently happened with "instructed action" depended on the context.
Measure	Make object specific measurement.	This code was used based on context. In most situations, it was used in (1) Cylinder 1 Task, Triangle 1 Task, Triangle 2 Task, and Volume Task, when resized the object, and (2) Cube Task, when put fingers on the faces, edges.
Resize	Make object bigger/smaller; change sides or angles.	This code was used when (1) changed the radius and/or height of the virtual cylinder in the cylinder tasks, or (2) changed the angle and/or length of the triangle in the triangle tasks, or (3) changed the size of virtual objects in the cube and volume tasks.
Move	Move objects to a different position without resizing.	This code was used when (1) moved any virtual object from one location to another, including vertical and horizontal movement, or (2) changed the orientation of the virtual object, such as Melinda changing the orientation of the cylinder in Cylinder 2 Task.

Miss	Miss picking up an object.	Cajun missed with her fingers when attempting to increase the height of the virtual object.

2. Hand Gestures

Representational /Iconic	Form a shape or object using only the hands.	This code was used when (1) fingers were used to represent a number, such as Vicky raising her five fingers to represent 6000, or (2) hands represented holding a shape, such as Cajun's pinching motion with her index finger and thumb to represent changing the height of the virtual cylinder, or (3) hand represented an abstract concept, such as Cathy making an "approximate" waving gesture with her right hand when she said, "it's hard to see exactly…."
Beat	Gestures that strengthen speech or provide emphasis on speech.	Riley beat with her hands when she said, "Ugh it's just like so hard to get like an even number."
Pointing	Point to someplace on the virtual object.	Vicky used her index finger to point to the virtual cylinder when explaining her understanding to Olivia.
Imitation	Imitate the facilitator or collaborator's gestures.	After the facilitator showed the gesture of using the index finger and thumb to cover the edges of the virtual cube, participants in Group 1 imitated this gesture.

3. Body (Head) Movements

Sweeping around gesture	Moving around the virtual object to get different perspectives.	This code was used when (1) moved around the virtual object horizontally to get different perspectives, or (2) moved around the virtual object vertically, such as Nancy lowering her body to observe the cylinder in Cylinder 2 Task.
Toward the virtual object	Walking or turning body towards the virtual object.	This code was used when (1) walked toward the virtual object or (2) turned her body toward the virtual object from other direction.
Toward the facilitator	Walking or turning body or head towards the facilitator.	When the facilitator raised questions, participants might turn their body to face the facilitator.
Toward other participants	Walking or turning body or head towards other participants.	When a participant was talking, other participants might turn their body to face the speaker.

4. Speech

Establishing Shared Understanding	Proposing a question, attempting to learn the perspective of others and trying to establish that what has been said is understood.	"So, make this angle right here, the top angle , smaller. Can you do that cause I can't. Move whatever angle to make this smaller".
Representing	Build a coherent mental representation of the problem and formulate hypotheses.	"……I guess it would look like a tunnel, but I'm not sure I could see depth if it were all in color, I think it would just look like an oblong oval."
Exploring	Explore and understand the problem space.	Val: "What is the one on the top, the arrow on the top do?"

| Planning | Develop a strategy or plan to solve the problem. | "……Let me try to change the radius back to like ten point three five." |
| Negotiating | Express agreement or disagreement and attempt to resolve conflicts when they arise. | The response could be either to the facilitator's question or other participants' point of view. The agreement or disagreement could be directly started with "yes" or "no" or implicit statements. |

After completing the labeling task, the labeled timeline was exported from V-note, and the of identifying interaction patterns was conducted in four steps. The first step was separating video records into episodes. According to van Dijk (1981), an episode is a semantic unit and represents some conceptual unity. A new episode may begin by a change in scene or participants, or the occurrence of a global event or action. Episodes are viewed as linked to each other, however, for building models of discourse processing. Due to the several layers of macrostructure, different episodes may be distinguished from one another within the same video record/story. Thus, the next task was to identify the most specific/detailed (i.e., low-level) episodes from the video records.

In this study, a low-level episode was the smallest analytic unit. As the study focused on collaborative learning, a minimal episode should include at least one interaction between at least two participants. An episode could begin with a new question proposed by the facilitator or with resuming the collaboration after a long pause. In general, episode length varies substantially. A continuous collaboration longer than 90 seconds or having a high-density of interactions might include too many actions/events to make sense of easily, which might impede finding meaningful patterns. In this type of situation, methods were used to separate the segment into two or more low-level episodes. For example, Figure 4 shows a labeled timeline for a complicated interaction process. In this timeline, Participant A introduced her idea of solving Triangle 1 as a conjecture to Participant B at the beginning, and then Participant B tries the idea by resizing the triangle. The labeled codes on the two sides of the yellow line in this video segment show different patterns. Thus, we separated this long collaborative segment into two low-level episodes based on their different interaction patterns.
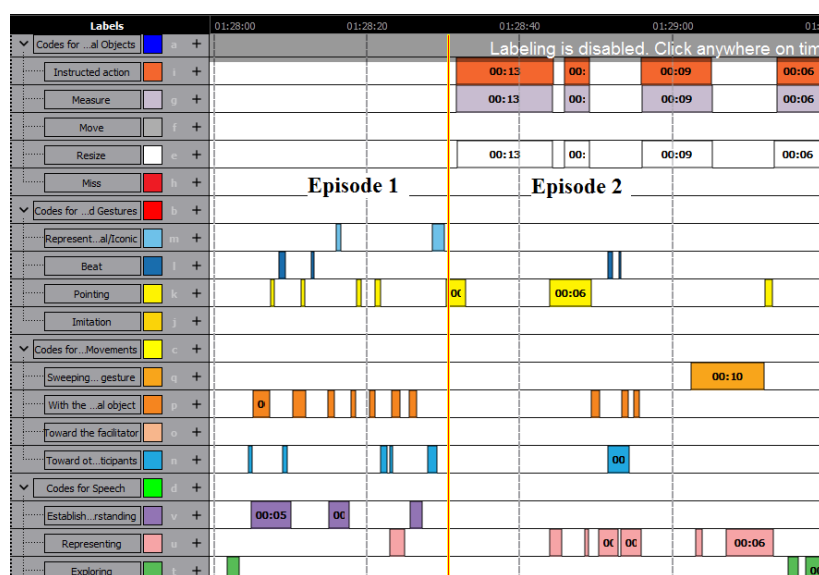

Figure 4. A complicated interaction timeline

Next, the relationship between each labeled action was established by adding lines on the labeled timeline. A no-arrow line means a simultaneous relationship. That is, a time overlap exists between the two actions, but it does not require these two actions to have the same start time or end time. An arrow line means a sequential relationship. The direction of the arrow indicates the temporal order of the two actions. This symbol is used in two situations. The first is two continuous actions within a single participant's action stream. The second is two actions belonging to different participants, in which these two actions should have a cause-effect relationship. If two actions appeared to accidentally happen in sequence, these two actions were not connected. An arrow line labeled "1" indicates the two actions have both simultaneous and sequential relationships, and the arrow represents the actions' sequence. Figure 5 shows an example in which Jill turns her virtual body towards other collaborators, then moves towards the virtual object (a cylinder). She then resizes the cylinder's height, an instructed action, and measures the change of the cylinder's volume. Melinda agrees with Jill's operation and sets a target for Jill to change the height of the cylinder ("Okay oh. Okay yeah. So, make it 22 cm"). The "J" and "M" denote each person's actions. A sequential relationship might exist between labeled actions in two different low-level episodes. In this situation, the respective scopes of relevant episodes (i.e., their start time and end time) would be adjusted so that the established relationship only exists between actions within the same low-level episode but not between ones belonging to different episodes.
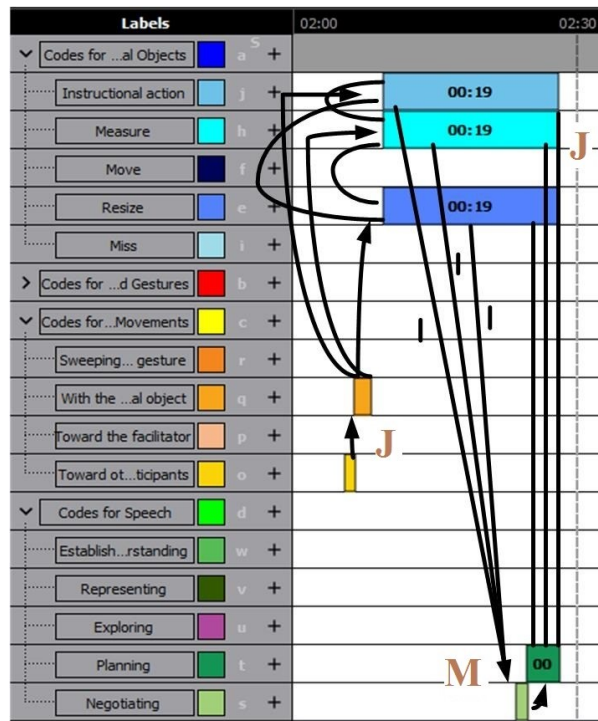

Figure 5. An example of establishing relationships on labels. "J" represents Jill and "M" represents Melinda

We recorded essential information about each episode. This information includes the basic topic, interactive patterns, simultaneous relationships, and sequential relationships. The basic topic is the main content of the episode. The interactive pattern is a short summary of relationship findings on the labeled timeline. Words in the summary are mainly from the codes defined in the codebook. The simultaneous relationships are a list of the paired action names that have established a

simultaneous relationship on the timeline. The sequential relationships are a list of the paired action names that have established a sequential relationship on the timeline.

When all low-level episodes had been developed, multimodal collaboration patterns were extracted by building high-level episodes for each conjecture task. Building the high-level episodes includes several rounds of combining episodes. Take a dataset having three levels of episodes, for example. First, continuous low-level episodes showing similar content/topics are combined into medium-level episodes. Then, continuous medium-level episodes (and remaining low-level episodes) are combined to develop high-level episodes. Table 4 shows a specific three-level episode dataset. These episodes are from the video records showing how participants (i.e., Jill, Melinda, and Nancy) answered the Cylinder 1 task by manipulating the cylinder and recalling the formula of cylinder volume. The first nine episodes are developed based on the timeline and represent the most specific details. Due to the content similarities among Episodes 1, 2, and 3, a medium-level episode (Episode 10) is built after combining these three episodes' content. Similarly, Episode 11 summarizes the content in Episodes 5, 6, and 7. The high-level episode (Episode 12) is a combination of these two medium-level episodes (Episode 10 and Episode 11) and the remaining low-level episodes (Episode 4, Episode 8, and Episode 9). The interactive pattern column in this table shows the pattern in each episode. In this case, both Episode 4 and Episode 5 show the pattern "Switch head between other participants and the virtual object," and both Episode 6 and Episode 10 show the pattern "Resize and speech simultaneously."

Table 4. An example of the episodes in Cylinder 1 Task

Episode No.	Start Time	End Time	Content	Interactive pattern
The Low-Level Episodes				
1	2:00	3:00	Jill changes the size of the cylinder. Melinda plans, and Jill represents her feeling about changing the size.	One resizes the height. Another one guides.
2	3:30	4:00	Jill changes the size of the cylinder. Melinda asks how to manipulate the cylinder.	One resizes the height. Another one explores with establishing shared understanding talk move.
3	4:00	5:30	Jill changes the size. Participants find the data showing the real-time height and size of the cylinder are different from each other and feel weird.	One resizes the height. Others raise an issue.
4	5:30	7:21	Participants report different volume data. Melinda changes the height of the cylinder. Still different data. Jill switches her head between the speaker and the virtual object.	(1) Resize the height after finding an issue. (2) Switch the head between other participants and the virtual object
	7:21	9:24	The facilitator writes the number on the blackboard.	No interaction
5	9:24	10:00	Melinda suggests Nancy try. Nancy does not know how to operate. Melinda explains. Jill switches her head between the speaker and the virtual object. Nancy starts to change the height.	(1) Resize the height after a suggestion. (2) Switch the head between the other participants and the virtual object.
6	10:00	10:39	Nancy changes the height. She claims she does not know what to do and loses her virtual hands.	Resize and speech simultaneously.
	10:39	11:00	The facilitator writes the number on the blackboard.	No interaction.
7	11:00	11:30	Nancy changes the radius of the cylinder.	
	11:30	14:10	The facilitator writes the number on the blackboard.	No interaction.
8	14:10	14:50	Melinda concludes that a significant difference exists between adding I cm height and adding 1 cm radius. Nancy agrees.	Representing talk move. Then negotiating talk move.
9	14:50	17:21	Participants discuss the conjecture combining the manipulation results and the formula.	One explores with establishing shared understanding talk move. Another uses negotiating talk move.

The Medium-Level Episodes

| 10 | 2:00 | 5:30 | Jill changes the size of cylinder. Others observe and communicate. | Resize action and speech simultaneously. |
| 11 | 9:24 | 14:10 | Nancy changes the height and then the radius of the cylinder. | |

The High-Level Episode

| 12 | 2:00 | 14:10 | Participants answer the conjecture by manipulating the cylinder in sequence. Then they use the cylinder volume formula to support their answer. | Action on objects first, then reasoning through language. |

Results and discussion

This study explores how learners use various embodied resources to collaboratively construct geometric knowledge in a shared VR environment. Using a novel multimodal data analysis method, we are investigating learners' interactive patterns at both the micro and macro levels. This section presents the data analysis results and answers the two sub-research questions. Additionally, we explain our findings and compare them to the results in prior relevant studies.

How do learners orchestrate modalities simultaneously or in sequence in shared VR environments?

A total of 108 episodes were developed (see Table 5). The average of a low-level episode is about 60 seconds (*Mean* = 62.7, *SD* = 38.6) with a broad range from 8 to 178 seconds. Although each group solved the same tasks, the specific solving processes could be different. Thus, the number of episodes is different between groups. These episodes include 1818 simultaneous relationships and 1723 sequential relationships between label instances. Tables 6 and 7 show the matrices of simultaneous and sequential relationships, respectively. The data in the matrices represent the number of established simultaneous and sequential relationships for each pair of codes. We did not use the amount of time for each pair of codes in the simultaneous relationship matrix. This is because the amount of co-occurrence time is more dependent on participants' verbal characteristics (e.g., speed of speech) and manipulative skills (e.g., proficiency with VR) and may not be sensitive to the characteristics of collaborative embodied learning.

Table 5. The number of episodes

Group	Low-Level	Medium-Levels	High-Level	Total
1*	23	4	5	32
2*	15	4	4	23
3	33	14	6	53
Total	71	22	15	108

Note. *Only one episode was developed for Cylinder 2 Task in Group 1 and Group 2, respectively. Only one episode was developed for Triangle 1 Task in Group 2. In these situations, it is not necessary to combine the single low-level episodes for building higher level episodes.

Table 6. The simultaneous relationship matrix

	IA	Msr	Rsz	Mv	Mss	Rp/Ic	Bt	Pnt	Imt	Swp	Obj	Fcl	Oth	ShU	Rpr	Exp	Pln	Ngt	Ttl
IA	0	106	103	32	4	4	7	11	0	12	37	5	8	35	86	14	37	75	576
Msr	106	0	97	10	0	3	7	10	0	10	32	5	7	26	74	10	36	71	504
Rsz	103	97	0	10	0	5	8	5	0	4	31	5	7	23	67	11	32	60	468
Mv	32	10	10	0	3	1	0	1	0	3	7	1	4	12	24	8	11	12	139
Mss	4	0	0	3	0	0	0	1	0	0	3	0	1	3	1	3	1	1	21
Rp/Ic	4	3	5	1	0	0	6	1	1	3	6	1	5	9	15	1	5	11	77
Bt	7	7	8	0	0	6	0	5	0	2	7	1	5	5	13	3	7	14	90

Pnt	11	10	5	1	1	1	5	0	0	3	10	2	2	17	24	5	13	19	129
Imt	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2
Swp	12	10	4	3	0	3	2	3	0	0	3	0	0	6	9	0	6	10	71
Obj	37	32	31	7	3	6	7	10	1	3	0	1	5	16	33	5	16	21	234
Fcl	5	5	5	1	0	1	1	2	0	0	1	0	0	3	5	4	0	8	41
Oth	8	7	7	4	1	5	5	2	0	0	5	0	0	7	18	7	7	14	97
ShU	35	26	23	12	3	9	5	17	0	6	16	3	7	0	8	10	7	6	193
Rpr	86	74	67	24	1	15	13	24	0	9	33	5	18	8	0	0	1	9	387
Exp	14	10	11	8	3	1	3	5	0	0	5	4	7	10	0	0	1	1	83
Pln	37	36	32	11	1	5	7	13	0	6	16	0	7	7	1	1	0	5	185
Ngt	75	71	60	12	1	11	14	19	0	10	21	8	14	6	9	1	5	0	337
Ttl	576	504	468	139	21	77	90	129	2	71	234	41	97	193	387	83	185	337	3634

Note. IA=Instructed action; Msr=Measure; Mv=Move; Mss=Miss;
Rp/Ic=Representational/Iconic; Bt=Beat; Pnt=Pointing; Imt=Imitation; Swp=Sweep gesture;
Obj=Toward the virtual object; Fcl=Toward the facilitator; Oth=Toward other participants;
ShU=Establishing shared understanding; Rpr=Representing; Exp=Exploring; Pln=Planning;
Ngt=Negotiating; Ttl=Total.

Table 7. The sequential relationship matrix

	IA	Msr	Rsz	Mv	Mss	Rp/Ic	Bt	Pnt	Imt	Swp	Obj	Fcl	Oth	ShU	Rpr	Exp	Pln	Ngt
IA	3	2	7	18	0	3	5	9	0	7	14	1	5	19	60	14	14	42
Msr	1	1	5	19	0	2	5	6	0	6	7	1	2	17	53	12	14	39
Rsz	4	3	1	21	0	3	5	5	0	2	10	1	2	14	42	11	13	27
Mv	26	25	31	4	1	1	1	0	0	2	7	0	5	6	12	1	1	4
Mss	1	1	1	1	0	0	0	1	0	0	2	0	0	1	2	2	0	1
Rp/Ic	1	0	0	1	0	1	1	3	1	2	6	1	1	2	2	0	0	2
Bt	1	0	0	2	0	2	1	3	0	0	5	1	2	0	2	0	0	1
Pnt	2	2	2	1	0	3	5	0	0	1	3	1	4	4	3	0	0	4
Imt	1	1	0	0	0	0	0	0	0	0	5	1	0	0	0	0	0	0
Swp	7	6	4	0	0	2	1	2	0	1	1	1	0	1	5	0	2	2
Obj	32	24	18	17	4	4	2	13	1	5	12	5	6	8	12	4	4	8
Fcl	0	0	0	0	0	2	0	0	3	0	16	0	7	0	1	1	0	0
Oth	0	0	0	1	0	6	0	2	0	0	44	4	1	4	7	0	0	5
ShU	6	3	3	5	0	2	2	4	0	0	9	2	7	10	14	1	10	37
Rpr	21	16	15	10	0	6	7	7	0	1	15	6	13	14	18	3	11	42
Exp	3	2	4	3	0	1	0	0	0	0	5	0	4	1	5	2	2	6
Pln	13	11	10	3	0	0	3	3	0	2	7	1	5	13	3	7	7	13
Ngt	15	11	9	5	0	5	3	9	0	1	8	2	10	11	17	1	15	22

Note. The row code represents the starting code, the column code represents the end code.
IA=Instructed action; Msr=Measure; Mv=Move; Mss=Miss; Rp/Ic=Representational/Iconic;

Bt=Beat; Pnt=Pointing; Imt=Imitation; Swp=Sweep gesture; Obj=Toward the virtual object; Fcl=Toward the facilitator; Oth=Toward other participants; ShU=Establishing shared understanding; Rpr=Representing; Exp=Exploring; Pln=Planning; Ngt=Negotiating.

Several themes can be observed from these tables. First, the relationships between labeled instances in Tables 6-7 highlight the intertwined relationship between performing actions on virtual objects and engaging in collaborative talk moves, with these two modalities occurring simultaneously or sequentially. This intertwined relationship can be explained through an extended and distributed cognition lens. From an extended cognition lens, actions on a virtual object can lead to a change in the properties of the virtual object (e.g., size, angle, and position); this change with the real-time data shown on the virtual object (e.g., size data and angle data) enhanced participants' ability to imagine and reason about geometric shapes, thus lowering cognitive task demands. Participants were then in turn more engaged in the representing and negotiating collaborative talk moves. From a distributed cognition lens, the virtual object was an external representation in the distributed cognitive system. This representation supported participants to maintain their coordination, express new ideas, and create shared knowledge by changing its states (e.g., size and position) (Ainsworth & Chounta, 2021). The representation's uniqueness disambiguated participants' references in collaboration. Thus, participants' verbal discussions and virtual actions were intertwined, gradually forming a resolution to the task (Chang et al., 2017).

Second, we see interesting patterns for body movements. The distinctive function of the virtual object in sharing ideas and building new knowledge can explain why facing the virtual object was the participants' usual body stance. The corresponding code "toward the virtual object" has high concurrent times with most codes in "collaborative speech" and in "action on virtual object." The "toward the virtual object" body movement code also shows high sequential incidences with most codes in "action on virtual object," as well as with the body movement codes "toward the facilitator" and "toward other participants." The facilitator and collaborators were able to talk and make gestures in order to attract a participant's attention, and the participant's head/body might then transitionally turn to face the facilitator or collaborators for communication. However, thinking and talking based on the virtual object's states (e.g., size and position) seemed critical for effectively solving conjectures. Thus, participants' bodies were then oriented toward the virtual object, and changes in this orientation seemed to signal an important and often purposeful shift in attention towards collaborators or the facilitator.

Third, gestures with the virtual hands have high *concurrent* times with collaborative speech. Specifically, gesture codes "representational/iconic" and "beat" have high concurrent times with collaborative speech codes "representing" and "negotiating." The gesture code "pointing" has high concurrent times with not only "representing" and "negotiating," but also "establishing shared understanding" and "planning." These findings are aligned with the claim that gestures often co-occur with speech (Novack & Goldin-Meadow, 2017) and the parallel use of speech and gesture form a joint embodied thinking process (McNeill & Duncan, 2000). Pointing gestures may occur when the speaker thinks or talks about a specific object in the environment (Hostetter & Alibali, 2019). Pointing gestures may orient the receiver toward contrasting spaces for a topic shift in a conversation (McNeill, 1992). In our study, the state of virtual objects contributes to sharing ideas. When participants communicated, they frequently pointed to the virtual object, no matter whether the virtual object was being manipulated or was in an idle state. This co-occurrence can explain why "pointing" also has high concurrent times with action codes "instructed action," "measure," and body movement code "toward the virtual object."

In contrast, the *sequential* relationship between gesture codes and codes involving actions on virtual objects or collaborative speech is weak in the sequential matrix. An explanation is that a gesture is overtly generated only when required constraints are satisfied (Hostetter & Alibali, 2008). The co-speech gestures may be used when the speaker emphasizes information (Church & Goldin-Meadow, 2017; Novack & Goldin-Meadow, 2017). These co-speech gestures are short and positioned "inside" most continuous speech segments (e.g., a speaker has a 10-second speech, the co-speech gesture only happens between the fifth and seventh seconds). In many situations, the cause-effect relationship is unclear between the gesture and the later speech, regardless of whether the speech and gesture are within a speaker or across speakers.

How do learners orchestrate modalities with collaborators in shared VR environments?
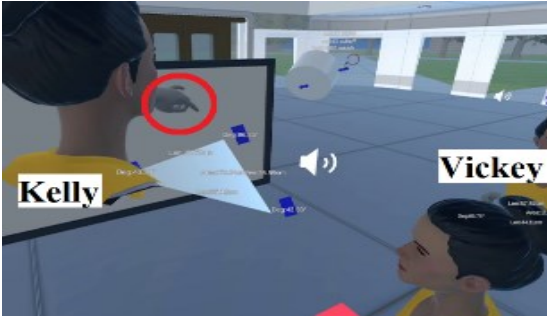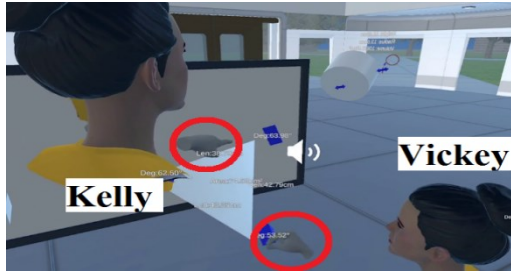
Five multimodality-interaction themes emerged after conducting interactive pattern analysis. These themes highlight how the modalities worked together to support participants in collaborative problem-solving in the shared VR environment, leveraging embodied resources. The first three themes are related to the different stages of collaborative problem-solving. These themes show that the interactions of multimodalities facilitated collaborators in presenting their ideas, communicate with their partners, and coordinate actions, effectively moving forward their reasoning to answer tasks. The fourth and fifth themes are about different strategies of collaborative problem-solving. A contrast of cases in these two themes shows that the multimodality-interaction pattern might affect whether the collaborators were able to respond to a task correctly.

Theme 1. Directed manipulation. Our first theme involved cases where Participant A directed Participant B with gestures or actions (e.g., pointing, instructed action, and measure), and then Participant B followed their advice to manipulate the virtual object. This kind of distribution of roles emerged as a key collaborative and embodied process in shared VR. Table 8 shows Kelly and Vickey taking these different roles in their collaboration. Kelly directed Vickey to resize the triangle to explore a conjecture about triangles, which stated that the largest side is always opposite from the biggest angle. In this scenario, the top corner of the virtual triangle had the largest angle, and its opposite size had the longest length. Kelly suggested to Vickey that she could make the top corner's angle smaller to explore this conjecture (Line 1). Kelly pointed to the top corner and performed a beat gesture (Line 3). Vickey turned her virtual body to face Kelly (Line 2), then turned back to the virtual object to resize and measure the triangle using Kelly's plan (Line 4). When Vickey resized the triangle, Kelly continuously directed Vickey using collaborative speech (Line 5). By manipulating the shape and collaborating via gestures, actions on objects, and speech, Kelly and Vickey were able to make valid generalizable observations about the relationship between linear and angular measures.

Participants' roles often emerged spontaneously or were negotiated by group members without interference by the facilitator during the shared VR learning task. Roles switches occurred frequently, allowing more consistent learning opportunities for all participants. Participant A might work as a director at the beginning of the session or task, and then change to a shape manipulator. At the same time, Participant B might begin as a manipulator and then switch to a director. Although less explicitly distributed behaviors (i.e., manipulating a shape based on your own idea while using a representing talk move to describe your thinking) sometimes happened, the participants often afterwards would appear aware of their dominant role, and would take a more secondary role as the collaboration continued. This activity structure that emerged from *shared VR*

is quite different from previous research showing that VR participants naturally adopt dominant roles when collaborating with desktop system participants or participants in the physical world (i.e., non-shared versions of VR; Kolomaznik et al., 2017; Slater et al., 2000). The difference could stem from the fact that our participants were in-service teachers with better collaboration skills than the participants in the two cited studies, primarily undergraduate students. However, it does point to a potential key interactional strength of collaborative VR environments, where all participants have the same learning information, materials, control, and immersion.

Table 8. An example of "directed manipulation"

1. Kelly: Move whatever angle except this angle to make this smaller. *[Kelly points to the top point]* *[Vickey turns to face Kelly and then turns back to face the triangle]*	
2. Vickey: So, this is the longest side. Do you want to make this one smaller? Is that what you're saying? 3. Kelly: Yeah, because that the longest length is always opposite of that, so you want to make is smaller to disprove it. So that is smaller and yeah. *[Kelly does a beat gesture, touches the bottom of the triangle from the left to the right and then points to the top corner]* *[Vickey moves the right bottom corner of the triangle around, making one angle close to the top angle]* 4. Vickey: Yeah, anytime I move the angle it is the longest side.	

| 5. Kelly: So that 71 is still the biggest one so you have to make it smaller.

[Kelly points to the top corner of the triangle]
[Vickey stretches the top corner out, makes an obtuse angle in the right bottom corner of the triangle] | |
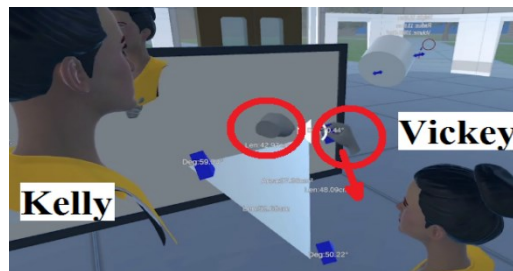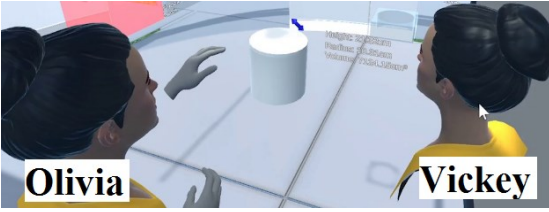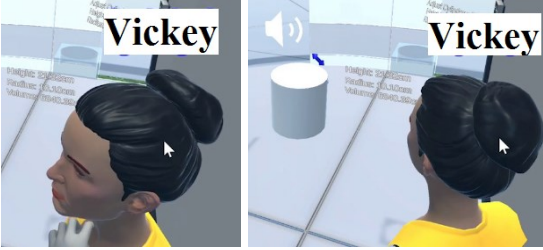
Theme 2. Switched attention. Our second theme involved cases where Participant A's body (head) moved between the virtual object and Participant B, accompanied with gestures or actions, when expressing their ideas to Participant B. Further, Participant B would simultaneously switch their body (head) to respond to Participant A. Table 9 is an example of this "switched attention" theme where a speaker's attention switches between different resources in the distributed cognition environment (i.e., virtual objects, interlocutors) to support group reasoning, which differs from "joint attention" that focuses on creating a single attentional focal point among interlocutors. Vickey explained to Olivia why she thought a 1-centimeter increase of the radius would change the volume of the cylinder more than a 1-centimeter increase of the height. Whenever there was a pause in Vickey's actions, Vickey switched her body direction (Lines 3-4, 6, and 7). When facing the cylinder, Vickey used gestures and speech simultaneously. She used a representational/iconic gesture when saying the original volume of the cylinder was 6000 (Line 3). Later, she used beat gestures when saying the new volume was caused by changing the cylinder's radius (Line 4). Additionally, Vickey had pointed to the cylinder twice - once at the beginning after she heard Olivia's request (Line 2), and then again when she started changing the height of the cylinder (Line 5). Olivia faced the virtual object at the beginning of this episode (Line 1), but when Vickey turned to face Olivia, Olivia changed her direction to face Vickey (Line 3). By discussing the resulting changes via speech, actions, and gestures, the collaborators were able to make conceptual progress towards the idea that radius increases the volume more because its measurement is squared and affects the area of the base.

During the attention switching process, the speaker established conditions to create joint attention so that the listener could follow the speaker's reasoning. In this example, Vickey used gestures and body movements to direct Olivia's attention. Olivia followed these directions to share a common point of reference with Vickey. In this conversation, the focus of joint attention changed. When Vickey faced and pointed to the virtual object, the virtual object served as the shared focus of attention. When Vickey faced Olivia, Vickey's representational/iconic gesture was the shared focus of attention. Previous literature indicated that gaze and the coordination between hand and eye are critical to coordinate switches of joint visual attention in the real world, used not only in the initiator's direction but also in the responder's confirmation (Reddy, 2011; Yu & Smith, 2013). In this shared VR environment, avatars' gazes and eye movements were unavailable. However, this limitation did not halt the establishment of joint attention between learners. This can be explained by people's capability to "compensate" for missing cues in a collaborative virtual environment (Roth et al., 2016; Steed & Schroeder, 2015). That is, other types of interaction can partially compensate for the absence of important behavior cues. In this study, body movement and pointing gestures compensated for gaze and eye movements to establish joint attention. In

addition, the distinctive function of the virtual object in sharing ideas and building new knowledge can also explain why Vicky turned to face the virtual cylinder several times in communication. This relates to the previously discussed high number of cases in the sequential matrix for the sequential body movements of "toward other participants" and "toward the virtual object."

Table 9. An example of "switched attention"

1. Olivia: I wasn't looking at the volume so can we try again? *[Olivia moves towards the cylinder, attempting to grab it]*	
[Vickey points to the cylinder] 2. Vickey: Yeah, I was saying that	
[Vickey and Olivia turn to face each other] 3. Vickey: when we the original was like 6000. *[Vickey uses the right figures to represent 6000]*	
[Vickey turns to face the cylinder] 4. Vickey: and I don't remember the exact. And when we changed the radius 1 centimeter higher it was 8200. *[Vickey beats with her hands]* *[Kelly is resizing the radius of the cylinder]*	

[Vickey points to the cylinder] 5. Vickey: And when we changed the height just and moved back to the original radius 10.3 and we changed the height by 1 centimeter,	
[Vickey turns to face Olivia] 6. Vickey: it was like 7000 something. *[Vickey turns to face the cylinder]* 7. Vickey: So, I……	

Theme 3. Responded action. Our third theme involved cases where Participant A initiated by acting on a virtual object (e.g., resize, move, measure). Participant B then turned her body to see Participant A's actions, then performed their own actions that responded to or built upon Participant A's actions. Table 10 is an example of Cathy and Riley adjusting the angles and length of their triangles to make the two triangles into a square. During this process, Cathy and Riley continued their adjustment after repeatedly observing the shape of their partner's triangle (Lines 1, 4, and 6). They also used speech to establish shared understanding and negotiate in this process (Lines 2, 3, 5, and 6). Through this activity, the participants negotiate mathematical meaning around the idea that only two isosceles right triangles can form a square, and that all legs of the triangles must be equal. They must achieve this meaning through embodied collaboration where they carefully coordinate their actions.

This interactive pattern may be somewhat similar to the worked example effect (Sweller, 2010), where the function of the partner's action and the object's shape were similar to a model solution. The observer looked at the partner's action and the object's shape, then moved her object to the next step. Because the observer only needed to consider each problem state and associated moves rather than an extensive range of possible moves in each step, this method was able to reduce the observer's cognitive load in solving an open-ended problem. However, different from the worked example effect, the model solution in the responded action pattern is temporary and imperfectly constructed. Thus, further language communication was needed to continue the progress after observation in each step.

From a distributed and embodied cognition perspective, this interactive pattern demonstrates that an embodied geometric reasoning activity can be distributed over multiple learners' actions and external media. The state of each triangle represented the knowledge of each participant. When the participant's partner observed the triangle state, this individual knowledge was shared and triggered the partner's cognitive processing. This pattern is similar to the alternating gestures described by Walkington et al. (2019) where learners observe and then build upon each other's gestures. The difference is that the virtual object, instead of the gesture, is the media used to make communication and collaboration in this responded-action pattern. Opportunities for learners to

experience these sequences of responded action in shared VR environments may be similarly essential to establishing effective embodied communication.
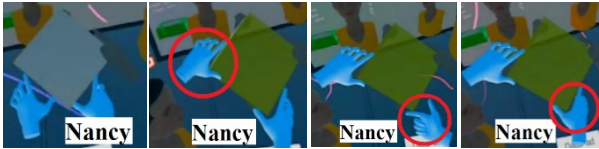
Table 10. An example of "responded action"

[Cathy turns to see Riley's triangle and then turns back to face her triangle]	
[Cathy resizes her triangle] 1. Cathy: Yours is 24. Let me change mine to 24.	
[Riley resizes her triangle, takes a moment to look at Cathy's triangle, and then continues adjusting her triangle] 2. Riley: Should we both change it to a 45, 45, 90? Maybe? 3. Cathy: Oh yeah that's good. 4. Riley: 45 oh.	
5. Cathy: But. Um. We want them all to be 24, right? Cause your sides are 24? *[Cathy looks at Riley's triangle, before adjusting her triangle once again]* 6. Riley: Let it 24, 24.	

Theme 4. Embodied action then discussion. Our fourth theme captures instances where participants started immediately with collaborative actions on the virtual objects (e.g., resize, instructed action, and measure) with or without accompanying speech/gestures. Afterwards, participants would answer the task based on the generated data or other references. Table 11 shows an example. Jill, Melinda, and Nancy worked together to solve the task: "Can you and your partner(s) use your index finger and thumbs to cover as many edges as possible?" Using instructed

actions on the object, they counted the number of edges they could cover (Line 1). Their actions directly allowed them to provide an answer to the task (Lines 2-3). Interestingly, though, this answer was not correct – given that each group member had 4 fingers to use (2 thumbs and 2 index fingers) and that a cube has 12 edges, they should have been able to cover all edges. However, the coordination required to cover all the edges is quite complex, given that thumbs are attached to index fingers and there are three people whose fingers need to be distributed across the shape. More initial discussion of the nature of the task might have better served this group's mathematical reasoning about polyhedra. This category is most significant when directly compared to our fifth and final category, so we now move to our fifth theme.
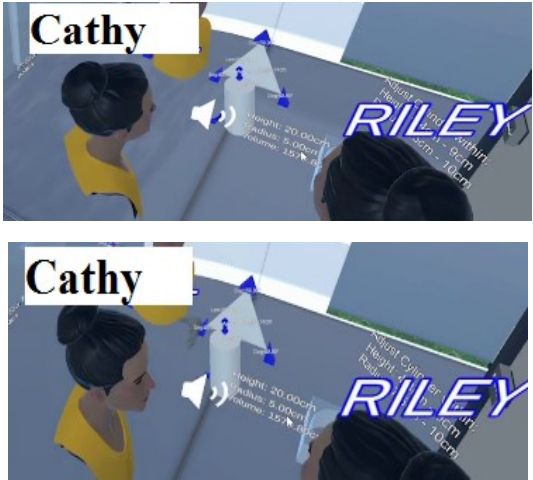
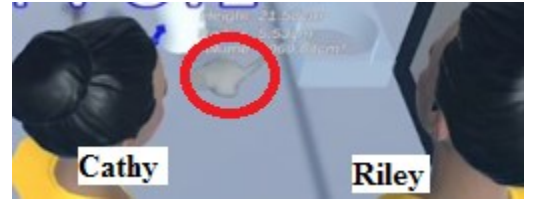Table 11. An example of "embodied action then discussion"

[Jill, Melinda, and Nancy quickly use their fingers to cover edges]	
[Nancy adjusted her fingers several times before her covering edges]	
1. Instructor: Okay so could three people do all of those? 2. Melinda: No. *[Melinda raises her hands. Her index finger and thumb show a right angle]*	
[Jill turns around to face Melinda] 3. Jill: Not with two hands.	

Theme 5. Discussion before embodied action. In our fifth and final theme, participants first collaboratively discussed a task using collaborative talk moves (e.g., negotiating and establishing shared understanding), representing their personal perspectives with gestures or actions (e.g., pointing and resizing). Only afterwards did participants implement their approaches with actions

on the virtual object to directly address the task (e.g., resize, instructed action, measure), accompanied with collaborative speech. Then they finally provided an answer to the task (e.g., presenting, negotiating). Table 12 shows an example where Cathy and Riley first both decided they disagreed with the conjecture that the volume of a cylinder changes by the same amount whether you increase the radius by 1 cm or increase the height by 1 cm (Lines 1-2), using collaborative speech to establish shared understanding and negotiate, and body/head movements to coordinate shared attention (Lines 1-2). Cathy then used Planning collaborative talk moves (e.g., Line 3) when they started engaging in actions on the virtual object (Lines 4-15). Then as the conversation progressed, they increasingly moved towards representation talk moves (Lines 10, 14). They transitioned to using gestures and additional actions on the object until they finally concluded with a representation talk move that contained no accompanying action (Line 17). By the end of the transcript, through exploring specific examples, they have jointly confirmed their initial insight about volume to better understand the different ways radius versus height affects a measure of volume.

Table 12. An example of "discussion before embodied action"

1. Cathy: If we change the radius one, it will increase the volume by a factor of squaring…. Greater than the height right, so no I don't agree, do you? *[Cathy moves head to look at partner, then turns head back to virtual object]* 2. Riley: No, I don't agree, mhmm.	
3. Cathy: Oh, so here's the radius, so let's change it, why don't we change it to one? No. Let's change it to 8…	
[Cathy resizes cylinder using right hand] 4. Riley: This is so… 5. Cathy: Okay 6.28, so it's… 6. Riley: Ok 7. Cathy: So, let's say 2000. 8. Cathy: And if I change it by 1, to like 7 something,	

[Cathy adjusts radius again] 9. Cathy: Then the volume changed to 3000, right? 10. Riley: Yeah, mhm, pretty close to 1000. 11. Cathy: And then, let's go back to whatever that was… *[Cathy adjusts radius again]*	
12. Cathy: it's hard to see exactly… 13. Riley: How does the height… *[Cathy makes "approximate" waving gesture with right hand]*	
14. Cathy: So, it's around 2000 volume, and it went up to 3000, but if the height goes up by 1… *[Cathy pulls on top of cylinder]* 15. Riley: Mhhmm.	
16. Cathy: It only goes up to 2059 *[Cathy points at measurements]* 17. Riley: Yeah, so the radius effects it more. 18. Cathy: Yeah.	

Themes 4 and 5 represent two different strategies for collaborative problem-solving in shared VR environments, where learners coordinate various modalities, including actions on virtual objects. Interestingly, all groups chose the strategy of "discussion before embodied action" to solve Triangle 1 Task and chose "embodied action then discussion" to solve tasks such as "Can three people use their hands cover all the cube faces?" in Cube Task. A possible explanation is that participants decision of whether or not to have an embodied discussion or engage in multimodal planning at the beginning was based on their perceived task difficulty. Solving the Cylinder or Triangle Tasks is more complex than the Cube Task. Participants were more inclined to discuss and plan how to solve the Cylinder and Triangle Tasks first, while participants might perceive that this step was unnecessary for the Cube Task. These two collaborative activity structures may have important implications for how tasks are designed in VR to initially allow or disallow (or direct or not direct) action-based manipulations and may also be an important signal of how the group perceives the task. It also may result from group composition, with some groups being more likely

to "jump into" the action on the virtual objects, while others prefer an initial planning process that may reduce later wheel-spinning.

Implications and significance

This study contributes to literature in three primary ways. First, it explores the function of 3D virtual objects in VR environments, namely dynamic models, in a CSCL learning activity, which extends theories of embodied and distributed/extended cognition. Second, it provides insights into the design of shared VR learning environments by identifying the important elements of role-switching, planning, task sequences, and realism with respect to additional embodied cues. Third, the proposed multimodal analytic approach showed strengths in analyzing how several modalities interact simultaneously and sequentially in collaborative learning; this method has the potential to apply in broader scenarios with further improvement. We now discuss each of these areas in turn in more detail, while discussing and building upon related literature in CSCL, virtual reality, and multimodal interaction. We then move to future considerations.

Embodied, distributed, and extended cognition

Both distributed cognition and extended cognition emphasize that when considering the human mind, we should not treat cognitive processing only within the scope of an individual's brain. Particularly, distributed cognition claims that external representations can be essential tools in knowledge sharing and coordination, and extended cognition indicates that external resources can complement the cognitive system of a hypothetically isolated brain. Individual-level volition and behaviors operate within this distributed system and are guided by group-level ideas embedded within a system that is mutually co-constructed by all of the constituent agents and resources. These group-guided behaviors, in turn, influence the cognitive processes of the participating agents through mechanisms such as transduction. A striking example of this is documented by Walkington et al. (2019), who found that collaborative gestures, in particular, can contribute to a joint activity with the properties of distributed and extended cognitive processing, in the case of mathematics education. The results of this study support these theories. Notably, the high concurrent times between "gesture" and "speech" during collaboration provide evidence to support the claim that gestures can promote a collaborative thinking process. Additionally, we extend these theories to dynamic virtual objects and actions on virtual objects in this study, using data from individuals, interactions between individuals, and group levels.

First, these dynamic virtual objects complemented an individual's cognitive processing. This point was well-illustrated by the intertwined relationship between "action on virtual object" and "speech." Their close relationship implies the value of manipulating virtual objects and the function of the real-time data on virtual objects in enhancing participants' ability to imagine and reason. Participants can then in turn be more engaged in representing and negotiating with others. Second, as a part of the distributed cognitive system, the state of these virtual objects also served as the carrier of shared knowledge across bodies, facilitating participants' communication. This phenomenon was illustrated by most themes found in this study. In particular, the example of Theme 2 "switched attention" shows Vicky frequently turned to face the virtual cylinder in her conversation, and Table 7 reports the high number of cases for the sequential body movements of "toward other participants" and "toward the virtual object." Finally, the state of these dynamic

virtual objects was able to advance the knowledge of a distributed cognitive system. A representative example is from Theme 4 "discussion then embodied action." The cube supported participants to answer the question, "Can you and your partner(s) use your index finger and thumbs to cover as many edges as possible" directly after they used the instructed action to cover the cube together. An individual would find it more challenging to answer without others' actions on the cube. Overall, as a novel type of artifact in CSCL, dynamic 3D virtual objects in shared VR environments benefit learners to explore mathematics in new ways and extend the appliable fields of existing cognitive theories.

Shared VR geometric learning environment design

The design elements of CSCL environments that best support learning are a key concern for researchers; as stated in Ludvigsen and Steier (2019), "concepts like scripts, scripting, prompts, orchestration and representations are used to describe how specific types of support can mediate participants' learning processes and outcomes" (p. 417). Support not only involves task characteristics, but also how teachers or facilitators interact, the roles and instructions given to students, and the way groups are composed.

Shared VR learning environments offer various advantages for learning with others, such as equal sharing of and access to multimodal resources and flexible communication channels. Depending on the group and task dynamics, a structured or semi-structured procedure may be needed to guarantee that these advantages actually lead to effective collaboration. Theme 1 "directed manipulation" implies the value of fluid roles in collaborative learning. Generally, the roles in collaborative learning can be either assigned by a teacher or they emerge or are negotiated by group members (Strijbos & de Laat, 2010). Participants may not have opportunities to collaborate effectively in shared VR environments without proper support. However, the fact that the learners using GSE were able to experience dynamic 3D objects together, coordinated in real time, all from the space of their own homes, is a remarkable accomplishment of VR. It also speaks to the rapidly-evolving set of locations in which CSCL activities occur as learning contexts and settings shift over time (Ludvigsen & Steier, 2019).

Theme 3 "responded action" shows that a complex task can be broken down into several steps that include embodied actions, and that each step may resemble a model solution as participants respond to each other's visible and shared reasoning. Participants can refer to the model solution to complete their task. This strategy has the potential to reduce participants' cognitive load, allow embodied actions to organically build on each other and be efficiently distributed in a jointly construed cognitive system, and enhance performance and learning. This kind of strategy can also be explicitly supported in the structure of VR tasks or in the ways in which learners are instructed to collaborate.

Themes 4 "embodied action and then discussion" and Theme 5 "discussion before embodied action" indicate whether participants stop and have a discussion or take time to plan at the beginning, before diving into complex sequences of interactions with the virtual object. These decisions may depend on their perceived task difficulty. Planning is crucial for collaboratively solving a complex problem with shared background knowledge about learning content (Shin et al., 2020). An absence of early planning may lead to investing more time later in problem-solving (Eichmann et al., 2019). However, a mismatch between learner perception of difficulty and the actual difficulty may occur. Thus, supports that provide structure for collaborative learning tasks

have the option of explicitly including or recommending a discussion or planning step at the beginning of the task.

The realism of a VR environment/avatars is not the only factor determining interpersonal interaction and learning performance. Due to the comparable capabilities between modalities, the realism of the environment/avatars not actually be needed in some VR collaborative learning tasks (Roth et al., 2016). In fact, a high degree of task-irrelevant realism may require learners to put more effort into comprehending these perceptions/mental representations in their working memory, thus weakening their performance (Ragan et al., 2015; Steed & Schroeder, 2015). Additionally, adding functions such as eye-tracking and facial expression tracking to the avatars requires an extra financial cost in terms of software and hardware. Thus, educators and researchers should evaluate the expected outcomes, software development time, and financial cost before determining the degree of realism needed in a shared VR environment. They should observe other ways learners can compensate for needs like shared/joint attentional focus, as they did in the current study by using hand gestures and body/head movements.

Multimodal analysis of collaborative learning

We used a novel multimodal analytic approach to identify participants' interactive patterns in the shared VR environment. This approach showed strengths in the process of multimodal data analysis, which provides implications for multimodality analysis in other CSCL environments. First, we found that the modalities to be analyzed should be precise, efficient, and rich. Multimodality can be any resources mobilized by participants for organizing their actions (Mondada, 2014). It may ultimately be too complex if we attend to all resources involved in the collaboration process. In this study, we focused on four modalities based on the design theory of the learning tasks and the characteristics of the shared VR environment. This focus clarified our research goals and simplified our analytic process. It also enabled us to look deeply and richly at how these four modalities were being leveraged.

Second, we found that visualizing coding/labeling was highly beneficial to interactive pattern finding. Previous studies mainly identified interactive patterns through coding transcripts on spreadsheets (e.g., Echeverria et al., 2019; Evans et al., 2011; Schroeder et al., 2006). We labeled video segments on the timeline in an exact time-coordinated manner and then used the labeled timeline to visually establish relationships between labeled instances. Compared to word-based transcripts, the labeled timeline offered an easier way to develop relationships and support us in finding interactive patterns faster and more accurately.

Third, multimodal data analysis can be conducted at both micro and macro levels. Descriptive analysis with counts of different codes and the way they occurred simultaneously or sequentially helped us identify how paired modalities interacted in collaborative learning. However, due to the complexity of the activities and elements of interaction, it was challenging to then extract how several modalities interacted together to create collaboration from our set of paired-modality interactive patterns. Thus, the macro-level data analysis used a method combining content analysis and interactive pattern analysis to accomplish this more complicated goal.

Nevertheless, this novel multimodal data analytic approach has limitations. One issue is that we initially could not distinguish whether the simultaneous or sequential relationship is established within a participant or between participants simply by a line or an arrow (see Figure 4). This is ultimately not a distinction we made within our pattern-funding in this paper. One possible solution is adding extra labels to identify the difference, such as attaching the initials of the participant's

name on the labeled instance. Another method is considering these relationships from the perspective of distributed cognition, and as not "belonging" to a single person. Our cognitive processing can be within our brain or across collaborators and between the individual and the material (Clark & Chalmers, 1998; Hutchins, 2000).

Future considerations

We also identify limitations that affect the results' generalization to broader populations and application scenarios. Specifically, there were only nine in-service mathematics teachers, three teachers per group, six geometry tasks, and one shared VR environment in the activity. Our purpose was to engage with a relatively "small" dataset in order to begin to understand the different ways in which modality channels could be coordinated and understood. More research on how embodied collaboration arises in VR is needed to continue to extend important activity structures and design considerations for collaborative learning with virtual bodies.

In the future, we plan to explore not only teachers but also students' collaboration in different shared extended reality environments (e.g., VR, AR, and mixed reality) with diverse mathematical tasks. We have designed new VR activities that target 7th and 8th grade geometry concepts, and that are intended to be used in groups of 3-4 middle school students working together. We are also implementing activities where high school students generate their own conjectures from dynamic AR holograms, rather than being presented with math questions with specific and well-defined answers. Finally, we are explicitly comparing the affordances of these virtual environments for learning about 2D versus 3D objects by examining if the dimensionality of the mathematical representations moderates the effectiveness of AR goggles compared to an iPad. These activities will enrich our understanding of the interactive patterns of collaborative learning in extended reality. Additionally, we are interested in exploring the possible relationship between students' interaction patterns and their learning outcomes in knowledge acquisition and problem solving. This further exploration can potentially promote the development of cognitive theories and extended reality-based learning environment design and is significant for the innovation of school mathematical courses.

Another question worth future investigating is what type of mathematical tasks are appropriate for justification and proof practices in shared VR environments. Our study shows that the function of actions on virtual objects is somewhat similar to simulated actions (gestures) in establishing justifications and proofs. That is, learners can directly get the answers for some conjectures through constructing meaningful, dynamic, virtual shapes in three dimensions. However, what makes action on virtual objects in shared VR environments different from gestures is that the shape's size, angle, or volume can be calculated immediately due to the computing technology embedded in the VR environment. Thus, learning in a GSE can help students establish justifications that utilize actions on virtual objects and thus move beyond language and gestures. However, it is important to acknowledge that VR equipment is still not yet widespread, and we do not yet have clear research guidance on whether it might be easier and more cost-effective to use laptops or tablets for mathematical objects that primarily operate in two dimensions. Thus, the field may need to develop a taxonomy of mathematical practices to advance the application of shared VR environments in geometry courses.

Furthermore, we plan to continuously improve the multimodal data analysis method proposed in this study. One challenge of applying this method in practice is the relatively heavy workload. For example, the reliability of the data analyses is determined by four aspects of fundamental work:

the code(s) labeled in an episode, the established relationship between labeled instances, the division of the low-level episodes, and the hierarchical episode structure. We only guaranteed an agreement on the codebook (including typical examples for each code) and the final identified pattern results in the study. In the future, an assisted software tool could be developed to support this kind of data analysis.

We mainly used qualitative methods in this study, but interactional patterns in larger datasets could certainly be analyzed using quantitative methods. The study results also inspired us to think more deeply about embedding quantitative assessments in similar VR activities. The possible assessments include measuring the learning outcome differences between participants in different roles and between groups using different collaborative problem-solving strategies or measuring the impact of avatar realism for embodied communication on participants' perception and performance. These assessments also can potentially guide the design of extended reality-based learning environments in the future.

In conclusion, this study contributes to developing relevant cognitive learning theories, designing shared VR geometric learning environments, and analyzing multimodal data in collaborative learning. It also identifies new directions for future research in these fields. This research can potentially benefit the spread of shared, extended reality technologies in mathematics education, which have the potential to enhance the quality of and change the nature of mathematical education in schools.

Acknowledgements

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H. & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education*, *5*(August), 1–29. https://doi.org/10.3389/feduc.2020.00147

Abrahamson, D. & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, *25*(2), 203–239. https://doi.org/10.1080/10508406.2016.1143370t

Ainsworth, S. E. & Chounta, I. A. (2021). The roles of representation in computer-supported collaborative learning. In *International Handbook of Computer-Supported Collaborative Learning* (pp. 353–369). Springer, Cham.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the learning sciences*, *21*(2), 247-286.

Andrews-Todd, J., Jackson, G. T. & Kurzum, C. (2019). Collaborative problem solving

assessment in an online mathematics task. *ETS Research Report Series*, *2019*(1), 1–7. https://doi.org/10.1002/ets2.12260

Benaquisto, C. L. (2008). Open coding. In *The SAGE Encyclopedia of Qualitative Research Methods* (p. 582). SAGE Publications, Inc.

Bieda, K. N., & Staples, M. (2020). Justification as an equity practice. *Mathematics Teacher: Learning and Teaching PK-12*, *113*(2), 102-108.

Bremig LLC. (2022). *V-note* (2.75). Bremig LLC. https://v-note.org/

Bujak, K. R., Radu, I., Catrambone, R., Macintyre, B., Zheng, R., & Golubski, G. (2013). A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*, *68*, 536-544.

Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Fan Chiang, S. H., Wen, C. T., Hwang, F. K., Wu, Y. T., Chao, P. Y., Lai, C. H., Wu, S. W., Chang, C. K. & Chen, W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers and Education*, *114*(300), 222–235. https://doi.org/10.1016/j.compedu.2017.07.008

Church, R. B. & Goldin-Meadow, S. (2017). So how does gesture function in speaking, communication, and thinking? In R. B. Church, M. W. Alibali & S. D. Kelly, (Eds.) *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 397–412). Philadelphia, PA: John Benjamins Publishing Company. https://doi.org/10.1075/gs.7.19chu

Cirillo, M., & Herbst, P. (2012). Moving toward more authentic proof practices in geometry. *The Mathematics Educator, 21*(2), 11-33.

Clark, A. (2012). Embodied, embedded, and extended cognition. *The Cambridge handbook of cognitive science*, 275-291.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420-464). New York: Macmillan.

Dehaene, S. (1997). *The Number Sense*. New York: Oxford University Press.

Deppermann, A. (2013). Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies*, *46*(1), 1–7.

van Dijk, T. A. (1981). Episodes as units of discourse analysis. *Analyzing Discourse: Text and Talk*, 177–195.

Dimmel J., Bock C. (2019) Dynamic Mathematical Figures with Immersive Spatial Displays: The Case of Handwaver. In: Aldon G., Trgalová J. (eds) *Technology in Mathematics Teaching. Mathematics Education in the Digital Era* (pp. 99-122), vol 13. Springer, Cham.

Echeverria, V., Martinez-Maldonado, R. & Shum, S. B. (2019). Towards collaboration translucence: Giving meaning to multimodal group data. *Conference on Human Factors in Computing Systems - Proceedings*, 1–16. https://doi.org/10.1145/3290605.3300269

Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L. & Naumann, J. (2019). The role of planning in complex problem solving. *Computers and Education*, 128, 1–12. https://doi.org/10.1016/j.compedu.2018.08.004

Ellenberg, J. (2021). *Shape: The hidden geometry of information, biology, strategy, democracy, and everything else.* New York : Penguin Press.

Ellis, A., Bieda, K., & Staples, M. (2022). Justification across the grade bands. K. N. Bieda et al. (eds.), *Conceptions and Consequences of Mathematical Argumentation, Justification, and Proof* (pp. 287-297). Cham: Springer International Publishing.

Enyedy, N. (2005). Inventing mapping: Creating cultural forms to solve collective problems. *Cognition and Instruction*, *23*(4), 427-466.

Evans, M. A., Feenstra, E., Ryon, E. & McNeill, D. (2011). A multimodal approach to coding discourse: Collaboration, distributed cognition, and geometric reasoning. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 253–278. https://doi.org/10.1007/s11412-011-9113-0

Fischer, U., Link, T., Cress, U., Nuerk, H. C. & Moeller, K. (2014). Math with the dance mat: On the benefits of embodied numerical training approaches. *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*, 149–166. https://doi.org/10.4324/9781315772639

Georgiou, Y. & Ioannou, A. (2019). Embodied learning in a digital world: A systematic review of empirical research in K-12 education. In *Learning in a digital world* (pp. 155–177). Springer Singapore. https://doi.org/10.1007/978-981-13-8265-9_8

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*(3), 389-399.

Harel, G. & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 6179, Issue 619, pp. 1–60). National Council of Teachers of Mathematics.

Healy, L. & Hoyles, C. (2000). *A study of proof conceptions in algebra. Journal for Research in Mathematics Education, 31*(4), 396–428. https://doi.org/10.2307/749651

Herbst, P. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics*, *49*(3), 283-312.

Hod, Y. & Twersky, D. (2020). Distributed spatial sense making on the augmented reality sandbox. *International Journal of Computer Supported Collaborative Learning, 15*(1), 115-141.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489–1522. https://doi.org/10.1016/s0378-2166(99)00096-x

Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B. & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin and Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, E. (2000). Distributed cognition. In *International Encyclopedia of the Social and Behavioral Sciences* (Vol. 138, Issue 1, pp. 1–10). Elsevier Science. http://www.slis.indiana.edu/faculty/yrogers/dist_cog/

Jewitt, C., Bezemer, J. & O'Halloran, K. (2016). Introducing multimodality. In *Routledge*. Routledge.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, *5*(81), 1–19. https://doi.org/10.3389/frobt.2018.00081

Johnson-Glenberg, M. C. & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications*, *2*(24). https://doi.org/10.1186/s41235-017-0060-9

Katic, E. K., Hmelo-Silver, C. E., & Weber, K. H. (2009). Material mediation: Tools and representations supporting collaborative problem-solving discourse. *International Journal of Teaching and Learning in Higher Education*, *21*(1), 13-24.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh* (Vol. 4). New York: Basic Books.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York, NY: Basic Books.

Lehrer, R., & Chazan, D. (1998/2012). *Designing learning environments for developing understanding of geometry and space*. Mahwah, NJ: Erlbaum.

Kolomaznik, M., Sullivan, M. & Vyvyan, K. (2017). Can virtual reality engage students with teamwork ? *International Journal of Innovation in Science and Mathematics Education*, *25*(4), 32–44.

Koschmann, T., & LeBaron, C. (2002). Learner articulation as interactional achievement: Studying the conversation of gesture. *Cognition and Instruction*, *20*(2), 249-282.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational researcher*, *42*(8), 445-452.

Ludvigsen, S., & Steier, R. (2019). Reflections and looking ahead for CSCL: Digital infrastructures, digital tools, and collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 415-423.

Marky, K., Müller, F., Funk, M., Geiß, A., Günther, S., Schmitz, M., Riemann, J. & Mühlhäuser, M. (2019). Teachyverse: Collaborative e-learning in virtual reality lecture halls. *ACM International Conference Proceeding Series*, 831–834. https://doi.org/10.1145/3340764.3344917

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago

Press.

Mcneill, D. & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. Mcneill (Ed.), *Language and Gesture* (Issue 1987, pp. 141–161). Cambridge University Press.

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, *65*, 137–156. https://doi.org/10.1016/j.pragma.2014.04.004

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, *20*(3), 336–366. https://doi.org/10.1111/josl.1_12177

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist*, *47*(2), 125–148.

Newen, A., De Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Norris, S. (2016) Concepts in multimodal discourse analysis with examples from video conferencing. *Yearbook of the Poznan Linguistic Meeting, 2*(1), 141-165. https://doi.org/10.1515/yplm-2016-0007.

Novack, M. A. & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin and Review*, *24*(3), 652–665. https://doi.org/10.3758/s13423-016-1145-z

Philippe, S., Souchet, A. D., Lameras, P., Petridis, P., Caporal, J., Coldeboeuf, G. & Duzan, H. (2020). Multimodal teaching, learning and training in virtual reality: A review and case study. *Virtual Reality and Intelligent Hardware*, *2*(5), 421–442. https://doi.org/10.1016/j.vrih.2020.07.008

Pidel, C. & Ackermann, P. (2020). Collaboration in virtual and augmented reality: A systematic overview. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12242 LNCS*, 141–156. https://doi.org/10.1007/978-3-030-58465-8_10

Pier, E. L., Walkington, C., Clinton, V. E., Boncoddo, R., Williams-Pierce, C., Alibali, M.A., & Nathan, M. J. (2019). Embodied Truths: How Dynamic Gesture and Transformational Speech Contribute to Mathematical Proof Practices. *Contemporary Educational Psychology*, *58*, 44-57.

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S. & Mcmahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, *21*(7), 794–807.

Reddy, V. (2011). A gaze at grips with me. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 137). MIT Press. https://doi.org/10.7551/mitpress/8841.003.0008

Rogers, Y. & Ellis, J. (1994). Distributed cognition: An alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*, *9*(2), 119–128. https://doi.org/10.1057/jit.1994.12

Roth, D., Lugrin, J. L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M. E. & Fuhrmann, A. (2016). Avatar realism and social interaction quality in virtual reality. *Proceedings - IEEE Virtual Reality*, *2016–July*, 277–278. https://doi.org/10.1109/VR.2016.7504761

Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Advancing the state of community knowledge. In International handbook of computer-supported collaborative learning (pp. 261-279). Springer, Cham.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I (Vol. 1)*. Cambridge university press.

Schroeder, R., Heldal, I. & Tromp, J. (2006). The usability of collaborative virtual environments and methods for the analysis of interaction. *Presence: Teleoperators and Virtual Environments*, *15*(6), 655–667. https://doi.org/10.1162/pres.15.6.655

Shelton, B. E., & Hedley, N. R. (2004). Exploring a cognitive basis for learning spatial relationships with augmented reality. *Technology, Instruction, Cognition and Learning*, *1*(4), 323.

Shin, Y., Kim, D. & Song, D. (2020). Types and timing of scaffolding to promote meaningful peer interaction and increase learning performance in computer-supported collaborative learning environments. *Journal of Educational Computing Research*, *58*(3), 640–661.

Slater, M., Sadagic, A., Usoh, M. & Schroeder, R. (2000). Small group behaviour in a virtual and real environment: A comparative study. *Presence*, *9*(1), 37–51.

Smith, C. P., King, B. & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *Journal of Mathematical Behavior*, *36*, 95–108. https://doi.org/10.1016/j.jmathb.2014.09.001

Stahl, G. (2015). Conceptualizing the intersubjective group. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 209-217.

Staples, M., Newton, J., Kosko, K., Conner, A., Cirillo, M., Bieda, K., Yopp, D., Zaslavasky, O., Hummer, J., Strachota, S., Singh, R., An, T., Going, T., & Zhuang, Y. (2017). Using Artifacts to Explore Conceptions and Consequences of Argumentation, Justification, and Proof. White Paper.

Steed, A. & Schroeder, R. (2015). Collaboration in immersive and non-immersive virtual environments. In *Immersed in Media: Telepresence Theory, Measurement and Technology* (pp. 263–282). https://doi.org/10.1007/978-3-319-10190-3

Strijbos, J. W. & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. *Computers in Human Behavior*, *26*(4), 495–505. https://doi.org/10.1016/j.chb.2009.08.014

Suh, J., & Moyer-Packenham, P. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, *26*(2), 155-173.

Sutton, J., Harris, C. B., Keil, P. G. & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the*

Cognitive Sciences, *9*(4), 521–560. https://doi.org/10.1007/s11097-010-9182-y

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Stylianides, Andreas J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3). 289-321.

Walkington, C., Chelule, G., Woods, D., & Nathan, M.J. (2019). Collaborative gesture as a case of extended mathematical cognition. *Journal of Mathematical Behavior, 55*, 1-20.

Walkington, C., Nathan, M.J., Wang, M., & Schenck, K. (2022). The effect of cognitive relevance of directed actions on mathematical reasoning. *Cognitive Science, 46*(9). https://doi.org/10.1111/cogs.13180

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71-95.

Weitnauer, E., Landy, D. & Ottmar, E. (2017). Graspable math: Towards dynamic algebra notations that support learners better than paper. *FTC 2016 - Proceedings of Future Technologies Conference*, *December*, 406–414. https://doi.org/10.1109/FTC.2016.7821641

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41-49.

Yu, C. & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0079659

Zheng, L., Xie, T. & Liu, G. (2018). Affordances of virtual reality for collaborative learning. *Proceedings - International Joint Conference on Information, Media and Engineering, ICIME 2018*, 6–10. https://doi.org/10.1109/ICIME.2018.00011

Zuckerman, O., & Gal-Oz, A. (2013). To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces. *International Journal of Human-Computer Studies*, *71*(7-8), 803-820.