

Transmodal Analysis of Embodied Collaborative Reasoning

Andres Felipe Zambrano¹[0000-0003-0692-1209], Ryan S. Baker²[0000-0002-3051-3232], Camille Giordano¹[0009-0001-2772-3079], and Stephen Hutt³[0000-0002-7041-7472]

¹ University of Pennsylvania, Philadelphia, PA, United States

² Adelaide University, Adelaide, SA, Australia

³ University of Minnesota, Minneapolis, MN, United States
afzambrano97@gmail.com

Abstract. Embodied collaborative reasoning involves both verbal interaction and bodily activity, as groups coordinate talk, perception, and material action to generate and evaluate evidence. However, prior research on collaborative problem solving has focused primarily on discourse, leaving less understood how embodied activity becomes part of shared reasoning. To examine how groups coordinate dialogic and embodied actions, this study uses Transmodal Ordered Network Analysis, comparing strategies between groups that reached correct versus incorrect solutions to a weight-estimation task. Results showed that both groups relied on bodily perception, but unsuccessful groups returned more heavily to individual sensations. In contrast, successful groups used a wider variety of embodied strategies, which helped structure perception more systematically and manage the limitations of their individual perception. Their dialogue was more focused on monitoring and establishing accepted findings than on merely sharing individual perceptions. These results highlight the importance of monitoring individual perceptions while extending embodied cognition to encompass the tools, materials, and interactions that support collaborative reasoning.

Keywords: Embodied Cognition, Collaborative Problem Solving, Transmodal Analysis, Ordered Network Analysis, Epistemic Network Analysis

1 Introduction

Collaborative reasoning and problem solving are increasingly recognized as essential competencies for 21st-century learning because they prepare learners to engage productively with complex challenges in dynamic contexts [1]. In physically embodied collaborative settings, these processes are often investigated through dialogic data, including think-aloud protocols and transcripts of students' interactions. Although these forms of evidence are important, focusing exclusively on them can overlook aspects of cognition that are tacit or difficult to express verbally. Learners do not reason only through speech and symbols; they also reason through embodied actions, defined here as physically situated gestures, movements, and manipulations of materials used to develop problem-solving strategies and make emerging ideas available to others [2, 3].

Bodily action and interaction with physical materials can contribute to the development of reasoning by helping learners notice relationships, test possibilities, reduce cognitive load, and coordinate shared attention around relevant features of a problem [4]. However, not all embodied activity contributes to learning in the same way [5]. Some movements may reflect exploration, uncertainty, offloading, or attempts to test an emerging idea, whereas others may help learners reorganize their understanding by changing what they perceive or how they relate different elements of a problem. Embodied actions can therefore function as evidence of a learner's current understanding and as a resource through which new understanding develops, making it valuable for identifying cognitive shifts or trouble spots before they are articulated in speech [6].

In collaborative settings, learners must not only reason individually but also make their reasoning available to others [7]. Embodied experiences, such as felt sensations, which are often initially private and subjective, must be communicated or transformed into shared evidence to support group reasoning. Thus, collaborative cognition involves coordinating individual bodily actions, perceptions, and reflections with verbal contributions [8]. Groups must develop common ground by deciding which embodied actions count as relevant evidence, how to interpret them, and how to integrate them with ongoing dialogic reasoning. This coordination presents a methodological challenge: analyses that focus only on discourse may miss important bodily and material aspects of the reasoning process, whereas analyses that examine movement or action in isolation may miss how those actions are interpreted and negotiated by the group. Addressing this challenge requires transmodal analytic approaches that examine how multiple modalities (e.g., speech, actions, and gestures) are coordinated within unfolding interaction [9, 10]. Such approaches can help identify not only which actions occur, but also how they are sequenced, repeated, and connected to the development of shared understanding.

Aligned with this challenge, the present study examines: **(RQ1)** How do groups coordinate dialogic and embodied actions during collaborative problem solving? and **(RQ2)** How do patterns of dialogic and embodied coordination differ between groups that reach correct versus incorrect solutions? This study uses Transmodal Analysis [9] and Ordered Network Analysis [11] to analyze conversations and actions during a weight-estimation task in which groups did not have access to measurement instruments. We examine how embodied actions are organized, calibrated, and accepted as shared evidence during collaborative problem solving. In doing so, this study contributes to research on embodied cognition and collaborative problem solving by showing how successful reasoning in physical collaborative tasks is characterized by the ways groups coordinate bodily perception, experimentation, and negotiation over time.

2 Related Work

2.1 Movement as a Resource for Reasoning

Embodied cognition literature argues that reasoning is distributed across perception, action, and the material environment rather than confined to internal symbolic manipulation [12]. This perspective has been influential in mathematics and science learning research, where bodily interaction can make magnitudes, spatial relations, and patterns

perceptually available [2]. Action and language can jointly support mathematical insight and proof, particularly in domains such as geometry that involve spatial and relational structures [13]. The role of bodily interaction is also evident in sensorimotor learning environments, where learners' movements are deliberately coordinated with perceptual feedback to support the development of mathematical ideas [3].

Kirsh and Maglio [5] distinguish between pragmatic actions, which move a person closer to completing a physical goal, and epistemic actions, which are performed to uncover information or simplify cognitive work that would otherwise be difficult to carry out mentally. This distinction is important because some movements may be incidental to reasoning, such as repositioning objects, adjusting posture, or managing task materials, whereas others may support comparison, calibration, hypothesis testing, or the reorganization of emerging ideas. Research on embodied cognition has therefore emphasized the importance of examining the cognitive relevance of specific movements, postures, and physical interactions with the environment rather than assuming that all movement contributes to reasoning in the same way [14].

2.2 Analyzing Collaborative Problem Solving Across Modalities

Successful collaboration depends on groups developing and sustaining shared understandings of task goals, emerging uncertainties, and possible next steps [7]. Much of the collaborative learning literature has examined this process through discourse analysis, treating dialogue as a central mechanism through which groups coordinate knowledge and make reasoning available to one another. However, reasoning in embodied collaborative tasks does not unfold through dialogue alone [15]. People coordinate talk with gaze, gesture, body posture, and object manipulation to produce coherent social action [16]. In educational contexts, prior work has shown how learners use their hands and bodies to display understanding during interaction [17], and how gestures in mathematics settings can extend, reinforce, or redirect shared mathematical ideas [8].

Embodied activity can also contribute information that is not initially available to other group members. Although gestures and material manipulations may be publicly observable and can shape how others perceive the task [16], bodily sensations are directly accessible only to the person experiencing them. For such experiences to become relevant to collective reasoning, they must therefore be made available through speech, visible action, or other forms of interaction [18]. Dialogue can organize and interpret embodied evidence, while embodied activity can introduce perceptual and material resources that influence the direction of group reasoning.

Studying these processes requires analytic approaches that capture how dialogue and embodied activity are coordinated over time. Multimodal learning analytics has emerged as one response to this challenge by expanding the forms of data used to study learning and providing richer accounts of learning processes in complex, interactive, and collaborative settings [10, 19]. Epistemic Network Analysis (ENA; [20]), Ordered Network Analysis (ONA; [11]), and Transmodal Analysis (TMA; [9]) offer useful tools to account for the different meanings, temporal structures, and levels of observability of different modalities. ENA has been used in collaborative learning research [10, 21] to model connections among coded events and examine how discourse and reasoning

are organized as networks of co-occurring practices. ONA extends this network-based approach by preserving the temporal ordering of coded events. TMA provides a framework for examining how different modalities interact over time without collapsing them into a single undifferentiated data stream (e.g., [22]), making it particularly suited to the study of collaborative embodied reasoning, as it can represent group strategies as temporally organized patterns of dialogic and embodied activity.

3 Methods

3.1 Data Context

This study used the publicly available Weights Task Dataset (WTD), a multimodal dataset released by researchers affiliated with four U.S. universities [23, 24]. The dataset includes video recordings, dialogue transcripts, gestures and collaborative problem-solving annotations, annotations of nonverbal behaviors (not available for all groups), and Azure Kinect body-pose data from 10 groups, each composed of three people working together at a shared table in a laboratory setting. Participants were adults recruited from the student body of Colorado State University. The sample was 80% male and 20% female; 60% Caucasian non-Hispanic, 30% Asian, and 10% Hispanic/Latino. IRB approval and informed consent were obtained for the analysis and public release of the data, including the video recordings of the experiment.

The activity consists of three tasks. In Task 1, participants were told the weight of one block and were asked to determine the weights of the remaining blocks using this known weight and a balance scale. In Task 2, groups were given a new mystery block and asked to infer its weight without using the scale, relying instead on embodied experimentation. Finally, in Task 3, groups were asked to infer the weight of another mystery block that was not physically present by recognizing the underlying pattern across the six weights they had observed in the previous two tasks. Overall, the goal across the three tasks was to infer unknown weights from observed patterns using external instruments and participants' own embodied sensations, while physically constructing the Fibonacci sequence, which served as the underlying pattern for the selected weights across tasks. The full dataset is available at: [10.5281/zenodo.10252341](https://zenodo.org/record/10252341).

The present study focused on Task 2 because it produced the greatest variation in outcomes and solution strategies. All 10 groups correctly completed Tasks 1 and 3 using relatively uniform approaches. In contrast, only 7 of the 10 groups correctly solved Task 2, while the remaining 3 produced incorrect responses and used different embodied and collaborative strategies. This variation made Task 2 an appropriate setting for examining differences in embodied and communicative processes across groups.

3.2 Coding Schemes

The original WTD provides multiple layers of qualitative codes [23, 24]. For Task 2 specifically, the available coded data include gesture abstract meaning representations (GAMR; [25]), collaborative problem-solving codes, and lift-and-put actions for each block. From these original coding layers, the present study focused on the collaborative

problem-solving codes because they indicate the group-level reasoning and interactional processes through which participants approached the task.

The original codebook organizes collaborative problem-solving codes hierarchically into three broad categories: constructive, negotiating, and relationship-maintaining. Constructive codes are further divided into sharing and establishing codes; negotiating codes are divided into responding and monitoring codes; and relationship-maintaining codes are divided into initiative and role-fulfillment codes. Each of these second-level codes is further divided into more specific subcodes, not used in our current analysis. In the original dataset, all coding categories were annotated by two coders with moderate interrater agreement ($\kappa = 0.62$) and subsequently adjudicated by an expert.

For the present analysis, we focused on the second-level codes within the constructive and negotiating categories: Sharing, Establishing, Responding, and Monitoring. These four codes allowed us to identify moments when ideas were proposed, questioned, or reflected upon, established as shared knowledge within the group, and developed through question-and-answer exchanges. Relationship-maintaining codes were not included because they were rare in the task, did not show substantial differences across the 10 groups, and were less directly relevant to tracing how participants' ideas and intuitions about the solution evolved during embodied testing. The final codebook for the dialog modalities is presented in Table 1.

Table 1. Dialog modality codebook used to analyze collaborative problem-solving interactions.

Code	Definition
Sharing	Contribution of ideas and perceptions to the group. Example. "Honestly, I still feel like we should go with one forty first, and if not, we should actually move to one thirty."
Establishing	Confirmation, clarification, or alignment of shared understanding with the groups often by restating the current state and checking for agreement. Example. "So right now this is one-twenty."
Monitoring	Checking current progress, evaluating whether a solution or action is working, identifying constraints, or deciding what should happen next. Example. "Actually, we can't have forty because we have this one."
Responding	Direct reply, or reaction to a partner's prior question, statement, or action. Example. S1: "How much was purple?" S2: "Purple was thirty."

Although the original dataset includes gesture codes annotated using the GAMR framework as deictic, iconic, emblematic, or metaphoric [25], data familiarization activities suggested that these categories did not fully capture the higher-level embodied strategies participants used to identify the unknown weight. Therefore, we developed a separate codebook (Table 2) focused on participants' embodied testing strategies. This codebook was developed after one annotator (the first author) reviewed all video recordings from the 10 groups, with refinements informed by discussions with a second annotator (third author). It distinguishes between instances in which participants relied primarily on lifting blocks by hand and comparing weights solely by feel, and instances in which they recognized the limitations of this perception and attempted to calibrate it. These calibration strategies included closing their eyes, switching blocks between hands, and testing their perceptions against blocks with known weights.

Table 2. Embodiment coding scheme used to analyze participants' embodied testing strategies.

Code	Definition
Hand-Based Block Lifting	Lifting or holding blocks using only the hands, without external objects, tools, improvised mechanisms, or additional physical setups.
Physical Experiments	Use of external objects or physical concepts to investigate the blocks without merely trusting on individual feeling.
External Information Representation	Use of written notes, annotations, drawings, tables, or other visible representations to record, organize, recall, or reference known information, prior results, constraints, or intermediate conclusions.
Perception Calibration	Adjustment of bodily perception to improve the accuracy of weight judgments. This includes closing eyes, changing grip, adjusting posture, or <u>testing perception using known weights</u> .
Weight Modification	Addition or reduction of weight involved in hand-based comparisons or physical experiments.
Blinded Testing	Testing and forming a judgment about weight while intentionally preventing other group members from seeing or being influenced by the test.
Tester Change	Shift in the person primarily handling, comparing, or testing the blocks.

We also coded more elaborate experimental strategies, such as testing the friction of each block, examining the center of mass, observing whether different weight configurations caused blocks to fall off the edge of the table where the task was conducted, or creating an improvised balance scale using materials available during the activity. In addition, we identified whether participants changed the person responsible for testing the weights, whether testing was conducted blindly to avoid biasing others' perceptions, and whether participants actively modified the weights or blocks during comparison. Finally, we coded whether participants used additional materials, such as a paper worksheet, to externalize their reasoning and record known information.

This codebook was applied at the event level. Two annotators reviewed the video recordings for all 10 groups. After discussing the code definitions and reaching a shared understanding of the coding scheme, they independently identified each embodied event performed by any participant, assigned the corresponding labels, and recorded the timestamp at the second level. Even before addressing disagreements, we observed Cohen's kappa values over 0.80 for all codes. Discrepancies in event labels and timestamps were then resolved by the first author after reviewing the videos again.

3.3 Transmodal Ordered Network Analysis

To examine how groups coordinated dialogic and embodied actions, we used ONA [11] within the TMA framework [9] employing the WebENA tool [26]. We used ONA because it allowed us to model both order and repetition of codes, which are particularly important for identifying differences in how groups coordinated talk and embodied action. For example, some groups engaged in dialogic discussion before or after physically testing a hypothesis, while others repeatedly modified the weights while keeping the same participant as the primary tester, or changed the tester while keeping the same weights to reach common ground about a specific comparison.

The analysis was conducted at the group level as the task was completed collaboratively and the coding schemes for both modalities were defined at that level. Groups

were further categorized by whether they successfully identified the weight of the mystery block (each group had two attempts). The two groups that identified the correct weight on the first attempt and the five groups that identified it correctly on the second attempt were grouped together, as they showed similar overall problem-solving patterns. Each group was treated as a single observation horizon [9], with no further segmentation. This decision was made because the task unfolded continuously (durations ranging from approximately 5 to 20 minutes) and the activity did not include additional breaks, phases, or formal subdivisions that would justify internal segmentation.

For both the dialogic and embodied modalities, we used a 60-second moving window to construct ordered networks. Because the coding was event-based and events were irregularly distributed over time, we selected a time-based window rather than one based on the number of prior coded lines. We tested alternative window sizes ranging from 30 to 300 seconds for each modality and found no substantive differences compared with the 60-second window. Therefore, the 60-second window was retained for the final analysis. We then compared the resulting group-level networks using a Mann–Whitney U test on the network centroids. This test determined whether the overall dialogic and embodied solution strategies of groups that solved the task correctly differed significantly from those of groups that did not.

4 Results

Fig. 1 and Table 3 present the TMA models comparing two groups: participants who produced correct solutions and those who produced incorrect solutions. The Mann–Whitney U test showed statistically significant differences in the overall use of dialogic and embodied strategies between the two groups ($p = 0.017$). The core strategy for all groups was using Hand-based Block Lifting to compare objects by feel (Fig. 2), although it was more central for groups that did not solve the task. This difference is reflected in the stronger Blocks Lifting $\rightarrow$ Blocks Lifting connection weights (CW) among unsuccessful groups (CW = 0.433 vs. 0.314). Groups that did not correctly identify the weight of the mystery block also focused more on changing the person responsible for lifting the blocks and forming judgments based on their own bodily perceptions, rather than engaging in other embodied actions, as shown by the stronger Blocks Lifting $\rightarrow$ Tester Change connection among unsuccessful groups (CW = 0.229 vs. 0.129).

Unsuccessful groups also engaged more heavily in repeated dialogic exchanges, particularly around sharing information (Sharing $\rightarrow$ Sharing, CW = 0.111 vs. 0.051). However, this sharing did not necessarily lead to establishing a common ground around their findings or conclusions. Instead, these groups were more likely to follow up their discussions about the mystery block by returning to the core action of lifting blocks as observed in their stronger connections for Sharing $\rightarrow$ Blocks Lifting (CW = 0.196 vs. 0.100), Responding $\rightarrow$ Blocks Lifting (CW = 0.217 vs. 0.133), and Monitoring $\rightarrow$ Blocks Lifting (CW = 0.233 vs. 0.168). Overall, unsuccessful groups appeared to rely more on verbal negotiation and repeated sensory comparison.

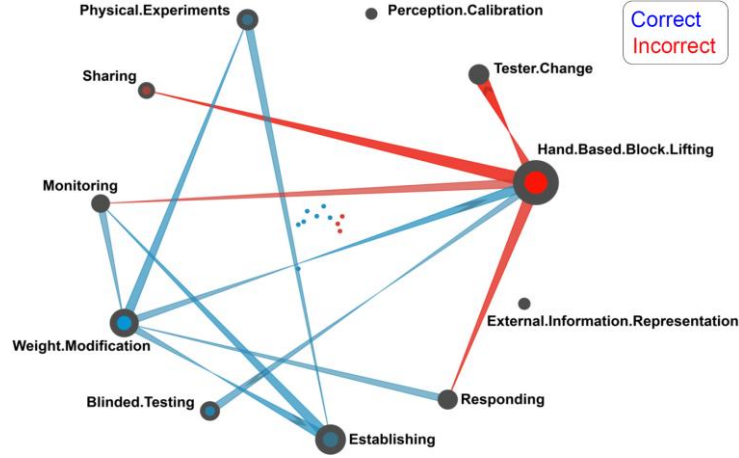


Fig. 1. TMA network model comparing groups with correct and incorrect solutions for the weight task. Edges with absolute group differences below 0.05 are omitted for visualization purposes.

Table 3. Connection weights (CW) from the TMA models comparing groups with correct and incorrect solutions for the weight task. Differences below 0.05 are omitted.

Connection	Correct	Incorrect	Difference
Hand-Based Lifting → Hand-Based Lifting	0.314	0.433	-0.119
Weight Modification → Weight Modification	0.141	0.029	0.112
Hand-Based Lifting → Tester Change	0.129	0.229	-0.101
Sharing → Hand-Based Lifting	0.100	0.196	-0.095
Tester Change → Hand-Based Lifting	0.122	0.210	-0.088
Responding → Hand-Based Lifting	0.133	0.217	-0.084
Blinded Testing → Blinded Testing	0.081	0.000	0.081
Weight Modification → Hand-Based Lifting	0.199	0.127	0.072
Weight Modification → Establishing	0.089	0.018	0.071
Monitoring → Establishing	0.142	0.072	0.070
Physical Experiments → Weight Modification	0.065	0.000	0.065
Monitoring → Hand-Based Lifting	0.168	0.233	-0.065
Sharing → Sharing	0.051	0.111	-0.060
Establishing → Physical Experiments	0.056	0.000	0.056
Hand-Based Lifting → Blinded Testing	0.055	0.000	0.055
Physical Experiments → Physical Experiments	0.054	0.000	0.054
Establishing → Weight Modification	0.115	0.062	0.054
Establishing → Establishing	0.162	0.110	0.053
Hand-Based Lifting → Weight Modification	0.180	0.128	0.052
Blinded Testing → Hand-Based Lifting	0.051	0.000	0.051
Weight Modification → Responding	0.069	0.018	0.051
Weight Modification → Monitoring	0.091	0.040	0.051

In contrast, groups that successfully solved the task performed a wider variety of embodied actions, and their dialogue was less focused on merely sharing information and more oriented toward monitoring their testing process and establishing what perceptions can be accepted as facts. One core strategy used by successful groups was

Blinded Testing (Fig. 3), in which each participant independently estimated the weight of the mystery block while the other two participants avoided observing the process. This allowed participants to form individual judgments before discussing their perceptions as a group, reducing the potential influence of others' trials. In some cases, groups then generated an initial guess by averaging the estimated weights across members. This strategy was observed only among groups that successfully identified the mystery weight, as indicated by the presence of Blinded Testing $\rightarrow$ Blinded Testing connections only in successful groups (CW = 0.081 vs. 0.000).

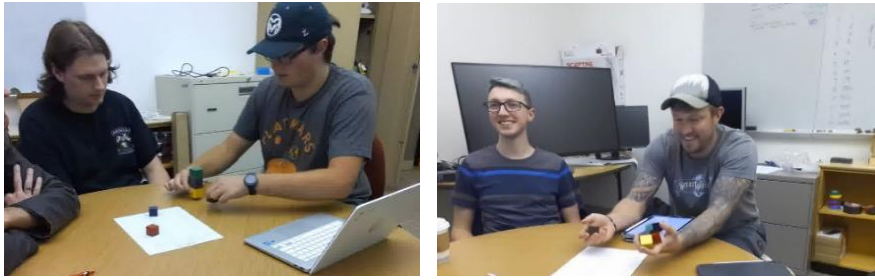


Fig. 2. Examples of groups using hand-based lifting to compare the weight of the mystery block with previously known weights. (All images publicly shared with permission in the original data set release).

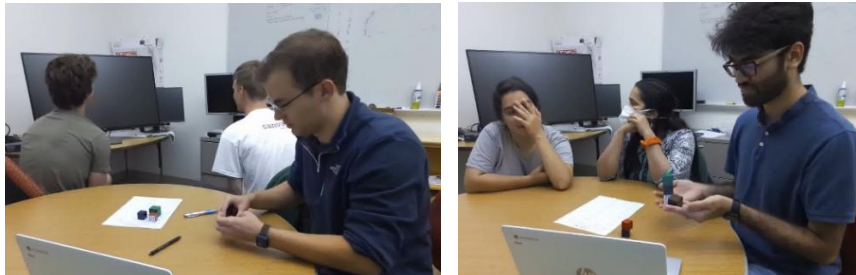


Fig. 3. Examples of blinded testing strategies.

Similarly, successful groups appeared to recognize the potential limitations of relying only on hand-based perception and therefore attempted additional Physical Experiments using objects available in the room. As shown in Fig. 4, these groups tried to create improvised scales using a tablet or a sheet of paper. Another group also used the edge of the table and the concept of center of mass to test the moment when enough weight was added to shift the blocks' center of mass and cause them to fall. As with Blinded Testing, these strategies were exclusively observed among successful groups, as reflected in the connection weight for Physical Experiment $\rightarrow$ Physical Experiment among successful groups compared with unsuccessful groups (CW = 0.054 vs. 0.000).

Another core distinction between the embodied actions followed by the groups was the role of Weight Modification. Although both successful and unsuccessful groups modified the weights while making hand-based comparisons or conducting additional

physical experiments, successful groups iterated more extensively on adding or removing weights during these comparisons. This pattern is reflected in the stronger Weight Modification $\rightarrow$ Weight Modification connection among successful groups ($CW = 0.141$ vs. 0.029). In contrast, unsuccessful groups tended to keep the same weight configurations and instead changed the person testing them, without engaging in similarly strong cycles of adding and removing weights as indicated by the stronger connection Lift Blocks $\rightarrow$ Change Tester previously mentioned.

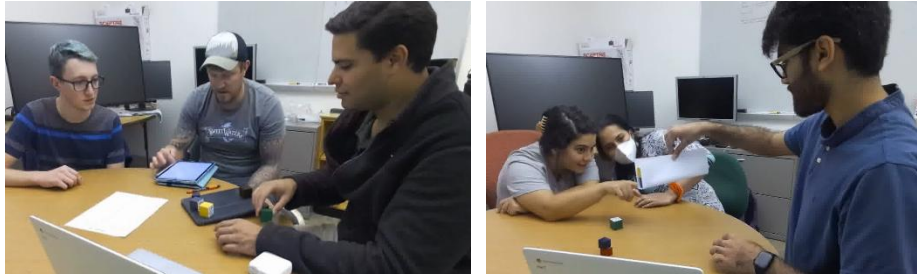


Fig. 4. Examples of physical experiments used to estimate the mystery block's weight.

Among successful groups, Weight Modification also tended to occur in response to physical experiments, as shown by the stronger Physical Experiment $\rightarrow$ Weight Modification connection among these groups ($CW = 0.065$ vs. 0.000). In addition, moments in which successful groups established accepted findings about the mystery block were strongly associated with iterative physical experiments involving weight modifications. This pattern was reflected in stronger connections between Establishing and Weight Modification among successful groups compared with unsuccessful groups ($CW = 0.115$ and 0.089 vs. 0.062 and 0.018 , respectively). Successful groups also tended to engage in Monitoring dialogue after making weight modifications ($CW = 0.091$ vs. 0.040), and this Monitoring, in turn, was more likely to precede dialogue in which groups established claims as accepted findings ($CW = 0.142$ vs. 0.072). An example of this cycle (Physical Experiment $\rightarrow$ Weight Modification $\rightarrow$ Monitoring $\rightarrow$ Establishing) can be observed in Group 3, which, after improvising a scale using a tablet and one of the blocks (Fig. 4) and testing multiple weights, had the following dialogue:

- P3: *It's leaning, still leaning that side, just remove one* (Label: Monitoring).
P1: *What ok so this one's fifty and thirty* (Labels: Responding and Establishing).
P3: *Fifty and Thirty* (Label: Establishing).
P2: *If you have, like, if you keep twenty [One of the blocks]* (Label: Monitoring).
[...Multiple Weight Modifications in the Scale...]
P2: *It should, it does still leaning* (Label: Monitoring).
P3: *So now it's leaning this way, if we put that it's leaning this way ninety* (Label: Establishing).
P2: *I think it's eighty now [80g is the correct solution]* (Label: Monitoring).
P1: *Yeah. That looks more even to me* (Labels: Responding and Establishing).

This excerpt illustrates how successful groups, after creating a physical setup for iterative testing, added and removed weights, monitored the resulting changes, and

gradually established certain observations as accepted findings until converging on the correct answer. Overall, these results suggest that successful groups did not rely solely on hand-based perception or the repeated sharing of information. Instead, they engaged in a wider range of embodied testing strategies while monitoring their findings and establishing conclusions from them.

5 Discussion

5.1 Implications for Embodied Collaborative Reasoning

This study examined how groups coordinated dialogic and embodied activity while solving a collaborative weight-estimation task. Although hand-based lifting was central across groups, it was more dominant among groups that did not solve the task. These groups repeatedly relied on individual sensations, often changing the person responsible for lifting the blocks and sharing their perceptions with the group. This suggests that they treated individual bodily sensations as the primary basis for judgment, which may have limited the range of evidence available for estimating the target weight. It may also have made it more difficult for the group to develop a shared body of evidence through which individual perceptions could be evaluated and established as collectively accepted findings. In contrast, successful groups moved beyond repeated sensory comparison by introducing additional embodied strategies, including blinded testing, physical experiments, and iterative testing of multiple weights. Blinded testing allowed participants to form independent judgments before discussion without being immediately influenced by others' perceptions, while physical experiments enabled groups to apply physical principles to estimate weight in ways that did not depend entirely on how the blocks felt in their hands. These strategies helped structure perception more systematically and, based on the groups' dialogue, appeared to support the evaluation of individual perceptions and the development of stronger common ground during the task.

These results may point to the importance of self-regulation [27], particularly reflection on the strategies employed and recognition of the limitations of bodily sensations as readily shareable evidence in embodied collaborative reasoning. Successful groups may have been more willing to tolerate uncertainty, double-check their perceptions, and revise their interpretations through additional testing. They may also have been better able to draw on prior knowledge and self-regulatory skills to propose and enact strategies, such as blinded testing or physical experiments, that enabled more precise estimates. Rather than treating an initial sensation as sufficient evidence, these groups created opportunities to compare, recalibrate, and validate their judgments.

By contrast, unsuccessful groups may have been more likely to settle on an answer prematurely or to have greater difficulty identifying alternative embodied strategies. These differences may have reflected several cognitive, metacognitive, and motivational factors. For example, some participants may have lacked sufficient knowledge of concepts such as center of mass or friction to devise physical experiments beyond hand-based perception, while others may have assumed that using additional materials was not permitted and therefore restricted their exploration of alternative strategies. Some participants may also have been less motivated to engage in iterative testing,

whereas others may have been overly confident in their initial perceptions and consequently tested fewer alternatives. Although these interpretations remain speculative, future research could examine them more directly by incorporating measures of prior knowledge, motivation, confidence, and self-regulation alongside analyses of dialogic and embodied actions.

From a theoretical perspective, these findings extend embodied cognition perspectives by highlighting the social and material conditions under which embodied activity supports collaborative reasoning. In this task, bodily perception was not automatically productive simply because learners physically engaged with the materials. Hand-based lifting generated perceptual information, but it remained potentially imprecise when treated as the single sensorimotor judgment. Embodied reasoning became more productive when groups used external materials and physical tools to structure perception and when they recognized potential sources of bias in their own judgments, such as differences in strength between their hands, changes in perception after seeing weights of different sizes, or the influence of observing others' judgments. Thus, embodied cognition in collaborative learning should not be understood only as individual sensorimotor experience; it also involves the material and social structuring of perception through tools, artifacts, the environment, and interaction with others.

Additionally, from a methodological perspective, this study demonstrates the value of Transmodal Analysis (TMA) and Ordered Network Analysis (ONA) for studying cognitive and collaborative processes as temporally unfolding multimodal phenomena. These methods are especially useful for examining how dialogue and embodied activity become connected over time, rather than treating discourse and material action as separate streams of data. In our case, the analyses highlighted cycles in which successful groups conducted physical experiments, modified weights, monitored the outcomes, and established observations as accepted findings. Although some multimodal analysis may require approaches that capture sequences involving more than two codes (e.g., Epistemic Association Rule Networks; see [28]), the pairwise connections examined here identified how groups transformed individual perceptions and material actions into shared evidence for collective inference, showing the value of TMA and ONA for research on embodied collaboration.

5.2 Limitations and Future Work

Although this study provides meaningful theoretical and empirical contributions, its findings remain specific to the task and context analyzed here. Future research should examine whether these patterns generalize across other tasks, domains, and populations. In addition, although task correctness was useful for comparing problem-solving trajectories, it does not fully capture the quality of students' conceptual understanding, the sophistication of their reasoning, or what they learned from the activity. Future studies could include richer measures of learning and conceptual understanding, such as pre- and post-assessments, explanations of the concepts underlying the task, or transfer tasks that evaluate whether students can apply what they learned in new contexts.

The present analysis also focused primarily on high-level embodied actions related to cognitive processes and solution strategies. However, embodiment plays an

important role in group communication as well [8]. In this dataset, most communication-related embodied actions appeared to be deictic, particularly pointing to or signaling objects with the fingers [23, 24]. Future research could examine a broader range of gestures and how they contribute to negotiation, coordination, and collaborative reasoning within groups.

Finally, coding both discourse and embodied action can be time-consuming, especially when conducting this type of research at scale with hundreds of participants. Future research could therefore explore the use of automated detectors for embodied and dialogic events, which would support the analysis of larger multimodal datasets and help scale this type of temporal analysis. Prior work has shown that both transcripts and videos can be coded using machine learning models [29, 30], although these approaches may still face limitations when applied to highly interpretive constructs or high-level strategies that go beyond signaling with fingers or body posture. Further research on automated coding could enable replication of this type of analysis with larger samples and more diverse contexts, supporting more generalizable claims and allowing researchers to investigate subtler and less frequent patterns of co-occurrence.

6 Conclusion

This study contributes to research on collaborative learning and embodied cognition by showing that successful embodied collaborative reasoning depends on how groups manage the imprecision of bodily perception and the potential influence of observing others' judgments. By examining the temporal connections among embodied and dialogic actions, this study highlights how groups can transform individual perceptions and material interactions into collective inference through monitoring, regulation, and the establishment of accepted findings. More broadly, the study shows the value of analyzing collaborative learning, using Transmodal Analysis, treating cognition as distributed across actions, tools, and the evolving social interaction of the group.

Acknowledgments. This study was supported by the National Science Foundation (Grant No. 2341410). The authors also thank Dr. Nathaniel Blanchard for recommending the use of the Weight-Task Dataset and for his contributions to its release.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Herlinawati, H., Marwa, M., Ismail, N., Junaidi, J., Liza, L.O., Situmorang, D.D.B.: The integration of 21st century skills in the curriculum of education. *Heliyon*. 10, e35148 (2024). <https://doi.org/10.1016/j.heliyon.2024.e35148>.
2. Alibali, M.W., Nathan, M.J.: Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *Journal of the Learning Sciences*. 21, 247–286 (2012). <https://doi.org/10.1080/10508406.2011.611446>.

3. Abrahamson, D., Bakker, A.: Making sense of movement in embodied design for mathematics learning. *Cognitive Research: Principles and Implications*. 1, 33 (2016). <https://doi.org/10.1186/s41235-016-0034-3>.
4. Goldin-Meadow, S., Nusbaum, H., Kelly, S.D., Wagner, S.: Explaining Math: Gesturing Lightens the Load. *Psychol Sci*. 12, 516–522 (2001). doi.org/10.1111/1467-9280.00395.
5. Kirsh, D., Maglio, P.: On distinguishing epistemic from pragmatic action. *Cognitive Science*. 18, 513–549 (1994). [https://doi.org/10.1016/0364-0213\(94\)90007-8](https://doi.org/10.1016/0364-0213(94)90007-8).
6. Alibali, M.W., Goldin-Meadow, S.: Gesture-Speech Mismatch and Mechanisms of Learning: What the Hands Reveal about a Child's State of Mind. *Cognitive Psychology*. 25, 468–523 (1993). <https://doi.org/10.1006/cogp.1993.1012>.
7. Roschelle, J., Teasley, S.D.: The Construction of Shared Knowledge in Collaborative Problem Solving. In: *Computer Supported Collaborative Learning*. pp. 69–97. Springer Berlin (1995).
8. Walkington, C., Chelule, G., Woods, D., Nathan, M.J.: Collaborative gesture as a case of extended mathematical cognition. *The Journal of Mathematical Behavior*. 55, 100683 (2019). <https://doi.org/10.1016/j.jmathb.2018.12.002>.
9. Shaffer, D.W., Wang, Y., Ruis, A.: Transmodal Analysis. *Learning Analytics*. 12, 271–292 (2025). <https://doi.org/10.18608/jla.2025.8423>.
10. Yan, L., Martinez-Maldonado, R., Swiecki, Z., Zhao, L., Li, X., Gašević, D.: Dissecting the temporal dynamics of embodied collaborative learning using multimodal learning analytics. *Journal of Educational Psychology*. 117, 106–133 (2025). <https://doi.org/10.1037/edu0000905>.
11. Tan, Y., Ruis, A.R., Marquart, C., Cai, Z., Knowles, M.A., Shaffer, D.W.: Ordered Network Analysis. In: *Advances in Quantitative Ethnography*. pp. 101–116. Springer Nature Switzerland (2023). doi.org/10.1007/978-3-031-31726-2_8.
12. Wilson, M.: Six views of embodied cognition. *Psychonomic Bulletin & Review*. 9, 625–636 (2002). <https://doi.org/10.3758/BF03196322>.
13. Nathan, M.J., Walkington, C.: Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications*. 2, 9 (2017). <https://doi.org/10.1186/s41235-016-0040-5>.
14. Walkington, C., Nathan, M.J., Wang, M., Schenck, K.: The Effect of Cognitive Relevance of Directed Actions on Mathematical Reasoning. *Cognitive Science*. 46, e13180 (2022). <https://doi.org/10.1111/cogs.13180>.
15. Sun, C., Shute, V.J., Stewart, A., Yonehiro, J., Duran, N., D'Mello, S.: Towards a generalized competency model of collaborative problem solving. *Computers & Education*. 143, 103672 (2020). <https://doi.org/10.1016/j.compedu.2019.103672>.
16. Streeck, J., Goodwin, C., LeBaron, C.: Embodied Interaction in the Material World: An Introduction. In: *Embodied Interaction: Language and Body in the Material World*. pp. 1–26. Cambridge University Press, Cambridge (2011).
17. Koschmann, T., LeBaron, C.: Learner Articulation as Interactional Achievement: Studying the Conversation of Gesture. *Cognition and Instruction*. 20, 249–282 (2002). https://doi.org/10.1207/S1532690XCI2002_4.

18. Rączaszek-Leonardi, J., Dębska, A., Sochanowicz, A.: Pooling the ground: understanding and coordination in collective sense making. *Frontiers in Psychology*. Volume 5-2014, (2014). <https://doi.org/10.3389/fpsyg.2014.01233>.
19. Mu, S., Cui, M., Huang, X.: Multimodal Data Fusion in Learning Analytics: A Systematic Review. *Sensors*. 20, (2020). <https://doi.org/10.3390/s20236856>.
20. Shaffer, D.W., Collier, W., Ruis, A.R.: A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*. 3, 9–45 (2016). doi.org/10.18608/jla.2016.33.3.
21. Csanadi, A., Eagan, B., Kollar, I., Shaffer, D.W., Fischer, F.: When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSDL research. *International Journal of Computer-Supported Collaborative Learning*. 13, 419–438 (2018). <https://doi.org/10.1007/s11412-018-9292-z>.
22. Borchers, C., Wang, Y., Karumbaiah, S., Ashiq, M., Shaffer, D.W., Aleven, V.: Revealing Networks: Understanding Effective Teacher Practices in AI-Supported Classrooms using Transmodal Ordered Network Analysis. In: *Proceedings of the 14th Learning Analytics and Knowledge Conference*. pp. 371–381. Association for Computing Machinery (2024). <https://doi.org/10.1145/3636555.3636892>.
23. Khebour, I., Brutti, R., Dey, I., Dickler, R., Sikes, K., Lai, K., Bradford, M., Cates, B., Hansen, P., Jung, C., Wisniewski, B., Terpstra, C., Hirshfield, L., Puntambekar, S., Blanchard, N., Pustejovsky, J., Krishnaswamy, N.: When Text and Speech are Not Enough: A Multimodal Dataset of Collaboration in a Situated Task. *Journal of Open Humanities Data*. (2024). <https://doi.org/10.5334/johd.168>.
24. Khebour, I.K., Lai, K., Bradford, M., Zhu, Y., Brutti, R.A., Tam, C., Tu, J., Ibarra, B.A., Blanchard, N., Krishnaswamy, N., Pustejovsky, J.: Common Ground Tracking in Multimodal Dialogue. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. pp. 3587–3602. ELRA and ICCL, Torino, Italia (2024).
25. Brutti, R., Donatelli, L., Lai, K., Pustejovsky, J.: Abstract Meaning Representation for Gesture. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. pp. 1576–1583. European Language Resources Association (2022).
26. Marquart, C.L., Hinojosa, C., Swiecki, Z., Eagan, B., Shaffer, D.W.: Epistemic network analysis (Version 1.7.0) [Software], <http://app.epistemicnetwork.org>, (2021).
27. Järvelä, S., Hadwin, A.F.: New Frontiers: Regulating Learning in CSDL. *Educational Psychologist*. 48, 25–39 (2013). doi.org/10.1080/00461520.2012.748006.
28. Zambrano, A.F., Baker, R.S., Mehta, S., Barany, A.: Epistemic Association Rule Networks: Incorporating Association Rule Mining into the Quantitative Ethnography Toolbox. In: Kim, Y.J. and Swiecki, Z. (eds.) *Advances in Quantitative Ethnography*. pp. 3–17. Springer Nature Switzerland (2024).
29. Liu, X., Zambrano, A.F., Baker, R.S., Barany, A., Ocumpaugh, J., Zhang, J., Pankiewicz, M., Nasiar, N., Wei, Z.: Qualitative Coding with GPT-4: Where it Works Better. *Learning Analytics*. 12, 169–185 (2025).
30. Whitehead, R., Nguyen, A., Järvelä, S.: Utilizing Multimodal Large Language Models for Video Analysis of Posture in Studying Collaborative Learning: A Case Study. *Learning Analytics*. 12, 186–200 (2025). doi.org/10.18608/jla.2025.8595.