JrnlID 11412_ArtID 9113_Proof# 1 - 09/03/2011

Computer-Supported Collaborative Learning
DOI 10.1007/s11412-011-9113-0

A multimodal approach to coding discourse: Collaboration, distributed cognition, and geometric reasoning

Michael A. Evans · Eliot Feenstra · Emily Ryon · David McNeill

Received: 16 September 2010 / Accepted: 3 March 2011

Abstract Our research aims to identify children's communicative strategies when faced with the task of solving a geometric puzzle in CSCL contexts. We investigated how to identify and trace *distributed cognition* in problem-solving interactions based on discursive cohesion to objects, participants, and prior discursive content, and geometric and cooperative concepts. We report on the development of a method of coding and representation of verbal and gestural content for multimodal interactional data and initial application of this framework to a microethnographic case study of two small groups of 7 and 8-year-old learners solving tangram manipulatives in physical and virtual desktop settings. We characterize the establishment of shared reference points as "coreferences" which cohere on object, para, and meta-levels through both gesture and speech. Our analysis foregrounds how participants establish common referential ground to facilitate collaborative problem solving with either computer-supported or physical puzzles. Using multimodal analysis and a theoretical framework we developed to study interactional dynamics, we identified patterns of focus, dominance, and coalition formation as they relate to coreferentiality on multiple levels. Initial findings indicate increased communication and cohesion to higher-level principles in the virtual tangram puzzle-solving setting. This work contributes to available models of multimodal analysis of distributed cognition using current manipulative technologies for early childhood mathematics education.

Keywords Collaborative learning · Communicative strategies · Coreferentiality · Distributed cognition · Early elementary mathematics · Gesture · Informal geometry · Multimodal analysis · Virtual manipulatives

This research was supported by National Science Foundation Grant IIS 0736151 awarded to Michael A. Evans.

M. A. Evans (✉)
Department of Learning Sciences and Technologies, Virginia Tech, 306 War Memorial Hall,
School of Education (0313), Blacksburg, VA 24060, USA
e-mail: mae@vt.edu

E. Feenstra · E. Ryon · D. McNeill
McNeill Speech & Gesture Lab, University of Chicago, Chicago, IL, USA

A multimodal approach to coding discourse: collaboration, distributed cognition, and geometric problem solving[1]

Our research aims to identify children's communicative strategies when faced with the task of solving a geometric puzzle (tangram manipulatives) in a group setting and their potential for exhibiting aspects of distributed cognition[2] in mathematics learning contexts. To facilitate the design and use of innovative strategies and technologies in the classroom, we have developed a "multimodal" system of coding and analyzing interaction to identify the ways children contribute to a knowledge-building interaction in a range of cognitive, perceptual, verbal, and nonverbal ways. Multi-dimensional coding schemes are by no means a novelty in CSCL research, but they are often not explicitly defined. As Strijbos and Stahl (2007, p.1-2) point out, what is needed is greater detail in the analytical methods and processes of multimodal techniques that will prove valuable to the community. These more detailed analyses are needed to understand the underlying mechanisms of group interaction.

Our research agenda is motivated by the construct of distributed, or group, cognition as a means to understand CSCL, advances in early childhood mathematics education, and the use of physical and virtual manipulative technologies. Though the extant CSCL literature contains noted references to group problem solving and joint construction of knowledge (Barron 2000; Kirsh 2009; Teasley and Roschelle 1993), along with collaboration of virtual math teams (Stahl 2006), our participant population is distinct in being much younger than those referenced in the cited studies. The rationale for investigating this population with specific manipulative technologies is provided below. Consequently, we follow Teasley and Roschelle (1993) closely in assuming that the basis of the framework of analysis is a relational, situated view of meaning: meanings are taken to be relations among situations and verbal or gestural actions (p. 1). We anticipate that the reported investigation, which provides an expanded view of gesture, builds on and adds to extant literature within the CSCL knowledge base.

For this article, we focus on the development of a theoretical frame and a multimodal analysis scheme to document distributed cognition (Hutchins 1995a, b; Hollan et al. 2000). In this article we will discuss our method of multimodal coding for analyzing co-located collaborative learning interactions in physical and virtual learning settings. By examining the children's speech, gesture, gaze and actions, we investigate the points of discursive cohesion that structure the children's collaborative reasoning throughout the problem solving process—objects, people, concepts, or mathematical principles referred to by multiple participants in the discourse. We identify these points of cohesion, or more simply repeated references to a single referent, as "coreferences" after McNeill et al. (2010). While most work on coreferential speech has focused on deictic pronouns and the relative transparency of the objects/persons they represent to the various interlocutors, in this study, we expand the concept of 'coreference' to include both verbal and non-verbal deixis.

[1] An earlier version of this work was presented at the "It's About Time" Workshop, Alpine Rendezvous 2009.

[2] In this paper, we take "distributed cognition" to be an emergent property of groups such that at the group level there's a thought process happening that's not fully instantiated by any one member of the group (S. Duncan, personal communication, August 21, 2010.) Shared orientation and focus and the mirroring of gesture or overlap of speech in the process of completing a task may demonstrate that individuals are "inhabiting the same state of cognitive being" (McNeill et al 2007). Distributed cognition is also an enduring interest of UC Santa Barbara linguist John Dubois. As Duncan put it, distributed cognition "gives the lie to the notion that we all function as message-lobbing monads" (personal communication, August 21, 2010). The notion of distributed cognition is a more general descriptor for the points of discursive cohesion, which we use as our primary units of analysis in this paper.

Based on the theory of coreferences and distributed cognition, we have developed a method of multimodal coding and analysis of collaboration to allow us to track nonverbal as well as verbal coreferences over time while attempting to retain the dynamism and narrative arc of the interaction. Our methodology is equally useful and applicable to settings where physical *or* virtual artifacts are used. By using multimodal coding software and controlled vocabularies to track coreferences, we are able to distill patterns and threads of communication and focus across related moments. The result is an expansive rubric for classifying how action, gesture, and speech relate to and build upon one another during the problem-solving process, which may be useful to other researchers seeking to find complex patterns in multimodal data.

We chose to focus on early elementary education (PreK through Grade 2 in the United States), which is becoming an increasingly important demographic for mathematics education research in the areas of problem solving and technology use. The data set reported here (several more sessions with additional tangram puzzles and triads are reported elsewhere) includes two groups of three 7 to 8-year-old children, one male group and one female, where each group was given a tangram puzzle to solve in two different settings: a physical set-up, using plastic pieces and a laminated reference sheet; and a virtual set-up, using a computer and mouse to maneuver the pieces into place on the monitor (see Fig. 1 for virtual tangram set up). We see our work contributing to mathematics education research on collaborative learning using virtual manipulatives, where attention to multimodal analysis is gaining a foothold (see Bjuland et al. 2008). We also offer a complementary method to those extant in the CSCL literature; see Cakir et al. (2010), Kershner et al. (2010), and Strijbos and Stahl (2007) for recent and relevant examples.

Theoretical frame

Distributed or group cognition

The construct of group cognition can be viewed as a powerful frame to analyze and describe learning in CSCL contexts. As Stahl (2006) suggests, distributed cognition places emphasis on meaning making as established in the interactive construction of referential networks. The discourse that arises from these interactions (for this paper

Fig. 1 Setup for virtual manipulatives trials and data collection. Arrangement was replicated for physical manipulatives

denoted by talk, gesture, gaze, and action) makes knowing visible thus permitting analysis and formulation of designs to further support CSCL. Consequently, distributed cognition as a focus for research and design guides our efforts as we attempt to better understand the development of mathematical thinking in PreK-2 (4–8 years old) students that will, eventually, lead to design specifications for technologies that support students' collaborative sense making. Our work is undoubtedly an extension of earlier work conducted by Stahl (2006) investigating virtual math teams where he examined co-located collaboration around the computer screen and used the analyses of interaction not only to understand group cognition, but to propose design specifications for technologies to better support argumentation and problem resolution (pp. 245–256).

Our analytical orientation, subsequently, would be more in line with a perspective formally referred to as distributed cognition (Hollan et al. 2000; Hutchins and Klausen 1998; Salomon 1993). In his work to formulate the construct of distributed cognition, Hutchins (1995a, b) described how computational aspects of navigation were distributed across a team of quartermasters and technology as they piloted an aircraft carrier off the southern coast of California. Hutchins' conclusion was that knowledge, work, and learning could be understood only if social interactions and cultural artifacts were taken into account. In essence, this view that brings together cognitive and sociocultural aspects of knowledge and work posits that "[t]he intellectual partnership that results from the distribution of cognitions across individuals or between individuals and cultural artifacts is a joint one; it cannot be attributed solely to one or another partner" (Salomon 1993, p. 112).

Collaborative learning and mathematical manipulatives

Papert (1980) suggests that physical and virtual objects play a central role in the knowledge construction process. He coined the term "objects-to-think-with" as an illustration of how objects in the world can become objects in the mind that help to construct, examine, and revise connections between old and new knowledge (Kafai 2006). Furthermore, as Figueria-Sampalo et al. (2009) noted:

> "…during the last few years, the number of cognitive conceptual tools based on constructivist principles has increased because they offer greater scope in achieving potential learning benefits than do traditional modes of instruction…The integration of new information and communication technologies has made transformations in teaching mathematics" (p. 485).

As pointed out by Tapper (2007), "manipulatives, like tangrams, help students build on prior knowledge and expand both their math content knowledge and their problem solving skills" (p. 11). Given that children have limited abilities to mentally transform shapes, activities allowing learners to experience and perform such transformations on physical and virtual objects can contribute to development and refinement of this ability (Clements et al. 2004). For example, while working on tasks involving the use of tangrams, learners must focus on translations (slides), reflections (flips) and rotations (turns) to make the pieces fit in the provided puzzle outlines, thus increasing their knowledge of transformational geometry in an explorative, constructive way (Moyer-Packenham et al. 2008).

Affordances of physical and virtual manipulative artifacts

Human activities and learning are profoundly influenced, or mediated, by the use of psychological and physical tools (Vygotsky 1978). Mediating artifacts include both externally oriented technical tools and internally oriented psychological tools or signs (Ares et al. 2009). In the educational realm, mediating artifacts means instructional strategies and technologies. Motivated by national standards that encourage appropriate use of technology to enhance mathematic learning (NCTM 2000), an increasing number of educational researchers now explore how to facilitate informal geometry via innovative instructional artifacts. Notable artifacts in mathematics instruction are virtual manipulatives (Moyer et al. 2002). Virtual manipulatives are interactive digital representations of physical counterparts (shapes, figures, and tiles) displayed on a computer screen and accessible over a communications network (Lee and Chen 2008; Moyer et al. 2002). Manipulative materials are objects designed to represent explicitly and concretely mathematical ideas that are abstract (Moyer 2001). The National Library of Virtual Manipulatives for Interactive Mathematics (Cannon et al. 2004) is an exemplar: a repository of Java applets that provide PreK-12 learners opportunities to engage in a range of open-ended exercises in basic mathematical categories, including geometry. Activities within the repository are constructed based on standards established by the National Council of Teachers of Mathematics (2000).

Despite documented advantages of virtual manipulatives (Clements 2000; Clements et al. 2004; Olkiun 2003; Reimer and Moyer 2005; Suh and Heo 2005), insufficient critical examination of their influence on mathematical thinking has been conducted. Although researchers may claim that “[t]he use of multiple representations can enhance the development of students’ abilities to think flexibly about mathematics topics” (Reimer and Moyer 2005), what has not been fully analyzed are the different ways physical and virtual manipulatives mediate mathematical inquiry and introduce particular affordances and constraints. A more nuanced understanding of the effects of manipulatives on children engaged in geometric sense making, particularly when working in small groups, would add to this literature base.

The necessity of a multimodal approach

In such a socially based arrangement, it is also relevant to consider the work of Bjuland et al. (2008), who find that the extent of mathematical communication occurring in a social context reaches beyond simple oral discourse:

> “…the pupils’ collaborative mathematical reasoning cannot be fully captured by only paying attention to what they write and what they say…Pupils’ gestures related to their use of reasoning strategies play a multifaceted role in developing mathematical reasoning in small groups” (p. 290).

The inclusion of the body in the act and process of knowing traces back to the phenomenological and epistemological work of Husserl (1931), Gehlen (1988), and Merleau-Ponty (1945). It is not because gesture is merely interesting but because it is in fact inseparable from language and meaning making that gesture, in conjunction with a wide range of other modalities, have come to be recognized as key elements in communication and conceptualization within science and mathematics (Roth 2001; Radford et al. 2009; McNeill 2009a; Kendon 2008). For example, the simple act of pointing allows students working collaboratively to focus the group’s attention to a particular portion of the puzzle; speaking the word, “Look!” means nothing if the rest of the team does not know where and at what they should be looking (Arzarello et al. 2009). However, gesture and speech do not always convey the same elements of meaning; they may be co-expressive if they capture the same idea, but each may express a different aspect of it (McNeill 2009a, ch 2). So, we see that the act of gesturing provides a context that spoken discourse alone is incapable of producing. A "multimodal" approach, as we have adopted in our analysis, aims to take into account the range of cognitive, physical, and perceptual resources that people utilize when working with mathematical ideas (Radford et al. 2009). The ability of students to effectively use gestures, as an additional form of communication, can be further refined through the implementation of both physical and virtual manipulatives, and whether these manipulative forms differently elicit gesture and other forms of communication. If differences were detected, then the design, implementation, and use should follow suit.

Analyzing collaboration in group problem solving with tangrams

Technical details of the research setup

We selected two groups of 7-8-year-old children: a group of three girls and a group of three boys. Although the students were similar in age, they differed in Test of Early Mathematical Ability (TEMA) based math competencies, grade level, gender, and experience with both tangram puzzles and cooperative mathematic problem solving (Table 1). A tangram puzzle is a dissection puzzle consisting of seven flat shapes that are put together to form a specific target shape, for example, a sailboat or bear. The objective of the puzzle is to complete this specific shape (given only in outline or silhouette) using all seven pieces, which may not overlap. Each group was given a tangram puzzle to solve in two settings: a physical setting with plastic pieces and a board, and a virtual setting in which the puzzle was on a computer screen and the children moved the pieces into place with a mouse. The virtual and physical sessions occurred on different days. To initiate each session, participants were reminded of the three basic rules for tangrams: 1) All seven pieces must be used; 2) No pieces can overlap; and 3) No pieces can extend beyond the lines of the target shape. At the beginning of the session, and at points when researchers detected frustration, the children were reminded to work together. If a group had not solved a puzzle after 5 min, the graduate research assistant provided a hint by placing a single piece in the correct location. Video footage was recorded from three angles to capture gestures and gaze of participants working in triads.

Multimodal analysis of group interaction

Collaboration is a difficult phenomenon to categorize and quantify because interactive behavior takes place in many different ways. Participants observe and respond to each other

Table 1 Descriptive demographic information of participants

Participant	Age	Gender	TEMA-3 grade equivalent	TEMA-3 age equivalent	TEMA-3 math ability score	TEMA-3 percentile ranking
Lauren	8-0	F	3.7	8-9	118	89
Mia	8-0	F	2.0	7-0	86	18
Rhonda	7–11	F	3.7	8–9	115	84
Steven	7–5	M	3.0	8-0	113	81
Jack	7–3	M	1.7	6–9	95	37
Adam	7–9	M	3.4	8-6	113	81

within different modes of discourse—verbal, gestural, and postural—and on different levels of discourse. We looked at an array of verbal and nonverbal indices of collaboration, including gaze, gesture, verbal utterances, and coordinated manipulation of physical or virtual puzzle pieces. The process of solving the puzzle is fueled by short-term cooperative action between two or all three participants, but if we watch and listen to longer sequences, repeating patterns begin to emerge:

> Individual utterances of certain semantic types (i.e., questions, answers, evaluations of answers) predictably follow one another to constitute an exchange. There are identifiable types of exchanges. These recur, recognizably for us and for the participants, not just for a while or among the same participants, but on different days, in different situations, and even in different classrooms in different schools. They constitute a cultural pattern or social semiotic *formation* (cf. Lemke 1995b) (Lemke 2000).

Q11

Patterns of types of exchanges can be seen across a number of timescales—from moment-to-moment toward much longer timescales of collaboration and knowledge construction such as across lessons or grades. Yet even within a single setting, the ebb and flow of communication and its continued integration into future interaction reveals detectable patterns of collaboration. To identify pivotal moments of collaboration, we sought to identify patterns in the structure of coreferences that drive the problem-solving forward.

To identify points of discursive cohesion, we looked for intervals of heightened interaction, where we looked at an array of indices for collaborative behavior. We categorized references using three *levels of discourse*: object-, para-, and meta-level coreferences. Using this system of coding, we were able to identify patterns of references that mark the introduction of new topics and periods of high and low productivity in puzzle solving.

Coreferences: Units of discursive cohesion

Our basic tenet is that discursive cohesion is necessary for successful group problem solving, and furthermore that interlocutors establish discursive cohesion via references to the same thing—objects, ideas, and other speakers. The way in which discourse coheres—how segments beyond individual utterances take form—can be observed in various ways, but we have found tracking coreferential chains that traverse verbal and nonverbal communication to be highly useful. A *reference* is an object or other meaning entity nominated in speech and/or indicated in gesture or action; a *coreferential chain* is a set, though not necessarily consecutive series, of linguistic and/or gestural nominations of the same referent that spans different speakers and links extended stretches of interaction. These coreferences can be categorized as follows:

- **Object-level coreferences** are references to an object or place in the physical world. (e.g., "this triangle," "here," or "oh look [*points at the right arm space*]")
- **Meta-level coreferences** are references to the discourse itself or to the problem solving process, including specific references to the computer program, and time limits. (e.g., "*that* wouldn't work" (where *that* represents a previous utterance) or "this triangle goes *next*" or "we *need to start over*" or "no *this way*" or "it's my *turn*" or "I *did it* whoo [*leans back and points at screen*]")
- **Para-level coreferences** are references to the participants themselves, the group, or emphasize a speaker's viewpoint. (e.g., "it's *your* turn," or "*I think*" or "I got you" or "*wait le-let me* do something real quick [*takes the mouse from R*]")

Many gestures or utterances contain multiple types of coreferences. Verbal object-meta coreferences are quite common (e.g. “that *(obj)* doesn’t fit *(meta)*”) as well as verbal para-object coreferences contained in the same utterance (e.g. “hey, look *(para)* here *(obj)*”). Often, gestures also rely on focusing attention on an object as well as another person or aspect of the past discourse. The overlapping and multifunctionality of verbal and nonverbal utterances support our general theory that coreferences tend to build upon one another, forming coreferential chains. These chains of cohesion comprise what we call simply *topics* in the discourse. Topics overlap—there may be a discourse about sharing the mouse overlapping with a discussion about the placement of a parallelogram—but they are nonetheless fairly distinct in the discourse. While we are able to recognize them by a “sandwiching” of para- and meta-level coreferences that show a shift in focus and the creation of a new “coalition” around the topic, our concept of topic units is observable by the trained unaided eye in the flow of communication among children. We observed that *the structure, length, and form of topic units differ significantly between the physical and virtual settings*.

Distinguishing further categories of meta- and object-level coreferences

For certain meta-coreferences, there need not be a previous utterance to serve as a meta-coreferent. Rather, there may be a rule or principle that provides the context for such an expression (often the rules included in the instructions given to the children before the start of the task). So far we have found that many utterances and gestures contain implicit references to geometric principles (fitting larger pieces in first, staying within the lines, particular properties of the pieces, etc.) as well as implicit references to principles governing collaborative problem-solving (like turn-taking, working together, etc.). Therefore, in coding, we distinguished between two types of meta-level coreferences: **mathematic** versus **project**. Differentiating between these two types of metacognition is useful in understanding the development of collaborative and problem-solving skills.

- “**Mathematic**” coreferences allude to geometric/ mathematic principles and the properties of puzzle pieces. (e.g.,. “*that fits*” or “it *keeps leaving that white space*”)
- “**Project**” coreferences adhere to collaborative problem-solving strategies or cooperation. (e.g.,. “let’s *start over*” or “*my turn goes next*”)

We think that both mathematic and project meta-commentary are key to the organization of distributed cognition but function in different ways within the discourse. Mathematic type meta-coreferences may be part of building skills in mathematical and geometric reasoning, as well as demonstrating an understanding of the geometric parameters of the task; whereas project type meta-coreferences cohere to the group dynamics and the implicit social rules of cooperation, collaboration, and step-by-step group problem solving. We also noted whether nonverbal object coreferences were based on **manipulation** (e.g. moving a square) or **gesture** (e.g. pointing to a square). While these are both clearly nonverbal object coreferences, they have different functions in the discourse.

Coalitions

We investigate how single coreferences (i.e. with a single referent) form *coreference chains*—multi-referential, multi-level accumulations of coreferential discourse. (In blind comparisons coders (Author B & Author C) agreed 83% on the level of coreference; after discussion this agreement was 96%.) Co-referential chains form when the participants align their focus on a single task within the greater aim of solving the puzzle (such as fitting in a certain piece, or filling in a troubling spot on the board) and where they seem especially responsive to each other. These chains shift levels (e.g. from meta- to para- to object-level) and do so particularly when the focus includes other participants. These shifts often signal the formation of a coalition, which often shows up as clusters of para- and object-level coreferences surrounding one or more meta-level coreferences. This makes sense in that a participant may join a coalition with a statement or action that recognizes the introduction of the new topic and indicates their allegiance to this theme (para-level), but in the course of the coalition they indicate the significance of the theme to the overall discourse (meta-level) (McNeill 2007; Cassell and McNeill 1991). The formation of a coalition (that is, a coreferential chain which sustains the focus of more than one participant) comprises a "topic" in the discourse. We are especially interested in how they form, who initiates them, whether they are characterized by agreement or disagreement, and whether or not all three children participate in them. Q12

Through identifying coalitions, we may be able to better understand pivotal moments in small-group collaboration. A given coreferential chain can span different speakers and can weave across different levels of discourse. By looking at who is speaking and participating, we can detect membership in a coalition. Figure 2a-c and Table 2 illustrate such a case. This section of discourse comes from the girls' physical setting and is one instance where we can observe the formation of a brief period where two participants (participants Mia and Lauren, shorthanded as "M" and "L" for coding purposes) are focused on the placement of the small triangle and then a third participant (Rhonda, or "R" in the coding and transcriptions) joins the coalition. M tries to fit the triangle as L watches, and then L advises her by reaching over and turning the triangle for her. M takes it back and fits it in and then R adjusts all of the pieces. The dialogue begins with the introduction of the triangle by M (Table 2, line 1) and is sandwiched by para- and meta-level comments that introduce the object, refer to the cooperation of the participants, and discuss the proper placement of the triangle. L's shift in focus in line 5 (Table 2) is begun with a para-level comment ("No") on M's action and then a meta-level reference to how the triangle should go. The topic ends with cohesion on the meta-level to the rule that all the pieces must fit within the lines and, as shown in Fig. 2c, M and R cooperate on this task while L watches. We can see object- and meta-level cohesion between M and L during this exchange and meta-level cognition shared among all three participants when R joins in on line 11 (Table 2). The patterning of coreferences and shared focus among all three participants mark this as a discrete topic unit and coalition of focus, initiated by M and joined by L and then R.

Temporality and methods of analysis for transcribing and coding discourse

The way researchers conceive of temporality in an interaction is largely a function of how it is represented in the transcript. The traditional transcript is arranged linearly, which sacrifices the accurate representation of overlapping and long-scale events for the purpose of readability. We also found that linear transcripts (such as the one above from the girls' physical setting) were not well suited to representing non-verbal behavior.

The alternative we chose was to work primarily in ELAN, a linguistic annotation software tool that was designed for the creation of text annotations for audio and video files of language use. Annotations are grouped on layers, in ELAN referred to as "tiers." Annotating activity on multiple tiers with a high degree of time accuracy allowed us to capture both sustained activity and overlapping events/levels. We designed a series of tiers

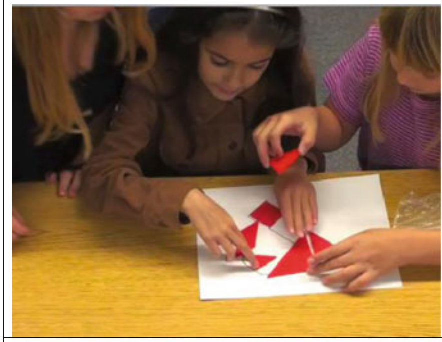
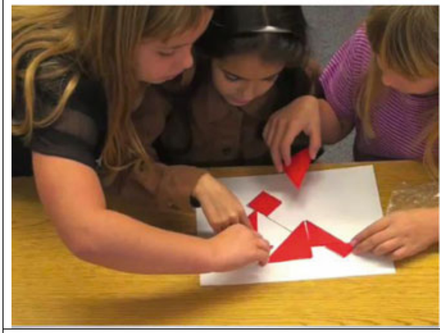
	Fig. 2a – M: "I think this triangle goes right here."
	Fig. 2b – L (to M): "No, turn it this way."
	Fig. 2c – R and M adjust the pieces that have slid out of place so that they fit within the lines.

Fig. 2 Mapping analyses to visual referents. **a**—M: "I think this triangle goes right here". **b**—L (to M): "No, turn it this way". **c**—R and M adjust the pieces that have slid out of place so that they fit within the lines

in ELAN for each type of index/action we wished to code independently, including gaze, gesture, and speech. We also created tiers to organize these actions and utterances into coreferences (types listed above), topic units, coalitions, and evidence of focus (see Figure 3 for the layout of our tiers.) ELAN creates the possibility for multimodal coding to be organized in a number of ways—for instance, we considered adding tiers such as "conflict" but later found we could sufficiently analyze the data using a more limited number of tiers.

Another advantage of ELAN is the possibility of using controlled vocabularies. We developed controlled vocabularies for several tiers/indices which allowed us to easily view

Table 2 Excerpt of transcription from girls' physical data, showing (in traditional linear transcription-style) the ordering of coreferences in a typical topic unit. Para- and meta-level coreferences are bolded because they are frequent indicators of the initiation and end of a topic unit or coalition

Line	**Participant**	Speech/Gesture/Action	Coreference type
1	**M**	I think this triangle goes right there	V Obj/**Para**/**Meta** (mathematic)
2	**M**	*(tries to place small triangle in space created from the last move)*	NV Obj (manipulation)
3	R	*(slides parallelogram back into place)*	NV Obj (mani.)
4	**M**	"um"	–
5	**L**	"no turn it this way"	V Obj/**Meta** (math)
6	**L**	*(turns small triangle)*	NV Obj (mani.)
7	R	*(picks up other large triangle and holds it)*	–
8	**M**	"I know this way"	V **Meta** (math)/**Para**
9	**M**	*(slides triangle from under L's fingers and fits it into place)*	NV Obj (mani.)/**Para**
10	**M**	"okay uh"	V **Meta** (project)
11	**R**	*(corrects the pieces that slid out of place)*	NV **Meta** (math)
12	**M**	"hmmmmm"	–

statistics for the coded sections—how often one participant looked at another or the computer screen, the types of object manipulation occurring, which students tended to participate in more meta-level coreferences, etc. These insights allowed us to make comparisons among the boys' and girls' groups, between physical and virtual settings, and among individuals (for instance, in comparison to their ranking in TEMA-tested mathematical skills).

However, while uniquely useful for multimodal coding, the ELAN interface is unfortunately also unwieldy and hard to read. It is difficult to see the linkage between verbal and nonverbal activity, even using the available statistics feature. There is no visual distinction between "organizing" tiers and basic transcription tiers, so we still needed a "linear" (Excel-style) transcript to serve as a sequential map of the interaction. Working with both Excel and ELAN, we were able to optimize our ability to visually represent both structural features and temporality.

While traditional transcripts, such as those that we later produced in an Excel spreadsheet, are useful for viewing/reading the overall arc of an encounter, for detailed verbal and nonverbal analysis, it is imperative that we were able to organize our observations to account for overlapping and long-scale actions. For a comparison of data representation in Excel and ELAN, see Figs. 4 and 5, another interaction from the girls' physical tangram puzzle-solving setting. Figure 4 is a screenshot from the first transcription, to ELAN, and Fig. 5 is a screenshot from the second transcription, to Excel. While Fig. 4, the ELAN version transcription of a section from the girls' physical puzzle-solving setting, shows a clear narrative arc and represents a "unit" of shared focus and singular topic, it is misleadingly sequential. Figure 5 (where time is the horizontal organizing variable at the top of the screen, down to 1/10 s divisions) is harder to decipher but the clear overlap of speech, gesture, and gaze as well as the lack of exact synchrony in the initiations of topic units and coalitions provides a more exacting image

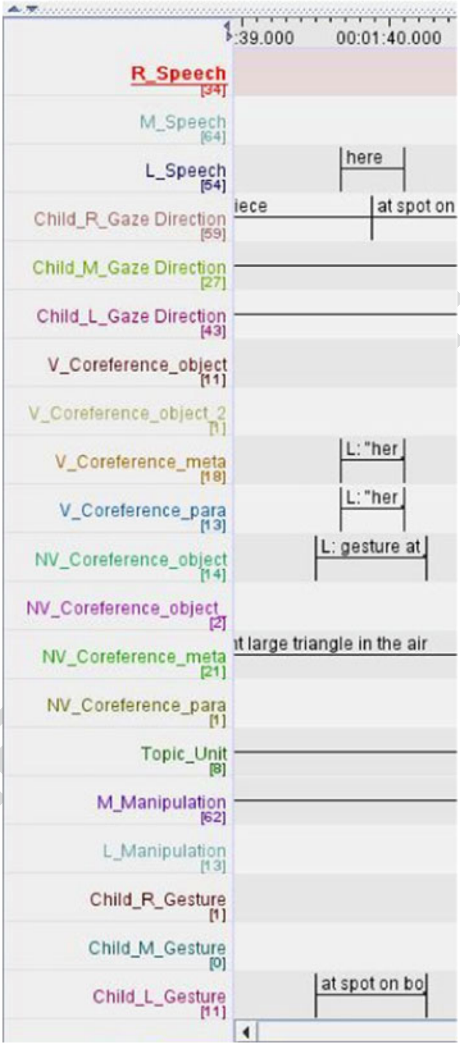


Fig. 3 ELAN tiers allow for multimodal coding which shows multisynchrony of gesture, speech, gaze and categorization of coreferences

of human interaction.[3] In the next section, we'll explore more closely how both ELAN and Excel are invaluable to our method of transcription and analysis as well as some of their limitations.

A more detailed example of multimodal transcription

While the gloss on Fig. 2a-c and Table 2 (in previous section titled "Coalitions") demonstrates briefly the formation of a coalition and topic unit, in this section we will

[3] For more on speech-gesture synchrony or lack of synchrony as an aspect of thought, that is, demonstrating "the joint presence of an idea unit in two [opposing] modes of semiosis [as] the form that human verbal thought takes," see McNeill 2009b. McNeill and others have written extensively about the "packaging" of linguistic categorical and imagistic components as a "growth point," the initial, dynamic pulse of thinking-while-speaking. Also see McNeill and Duncan 2000 for more on growth points.

Time	Participant	Speech/Action	Gaze			Medium of cohesion	Focus (seconds)	# of part.s with common focus	Initiator of topic unit	Initiator of coalition
1:34	M	"(()) this-- I think that this triangle goes next"	B	B	M/m	V meta/para		3		
	M	(takes large triangle that R was holding)	M/m	S.O.O.F	M/m	NV obj/para (mani)		3		
	M	(tries to place large triangle in center)	M/m	B	M/m	NV obj (mani)		3		
1:39	L	"here"	M/m	M/m	M/m	V para		3		
	L	(gestures towards board)	M/m	M/m	B	NV obj/para (mani)		3		
	R	(points to corner in center)	B	R/g	B	NV obj (gest)		3		
	L	(points to same corner)	B	B	B	NV obj (gest)		3		
1:41	R	"look put the tip of it right--"	B	B	B	V obj/para		3		
	R	(takes large triangle from M and tries to put it in said space)	R/m	R/m	R/m	NV obj/para (mani)	M takes large triangle from R and tries to fit it	3		
1:44	M	"wait I think--"	R/m	R/m	R/m	V para	into the center. Both L and R point to a spot and	3		
1:44	R	"no"	R/m	R/m	R/m	V meta	advise M to put the tip of the triangle into a	3		
	M	(takes back large triangle from R)	M/m	M/m	M/m	NV obj/para (mani)	corner. R takes the piece from M and tries to fit it	3		
	R	(adjusts pieces that have slid out f the lines)	M/m	M/m	R/m		in. R realizes it won't fit and M takes it back	3		
	M	(clears another piece out of the way)	M/m	M/m	M/m	NV obj (gest)	from her. L looks on. M switches the orientation	3		
1:45	M	"I think it goes like this"	M/m	M/m	M/m	V obj/para	of the large triangles to that their hypotenuses	3		
	M	(places large triangles so that the hypotenuses face each other)	M/m	M/m	M/m	NV obj (mani)	face each other	3	M	L and R

Fig. 4 Data from girls' physical setting represented in linear transcript using Excel, allowing researcher to observe temporal flow of coreferences relative to formations of topic units, coalitions, and shared focus **Q15**

demonstrate how the interpretation of a specific piece of transcript, using this coding method, allows us to keep moments and dynamism alive in the transcript and enable researchers to identify patterns and coalitions from the transcript.

The first step in transcribing and analyzing this interaction is the basic transcription of speech into Praat. We use Praat, an open-source multifunctional program for analysing, synthesizing, and manipulating speech, to transcribe speech. Using Praat allows us to annotate speech with a very high degree of temporal accuracy; we then import the Praat textgrid file into ELAN, where we annotate gaze, and gesture from video; in Fig. 5, these three categories of annotation are captured for each participant in tiers 1–3 (speech of each participant, R, M, and L), 4–6 (gaze, one tier per participant) and 16–21 (gesture and

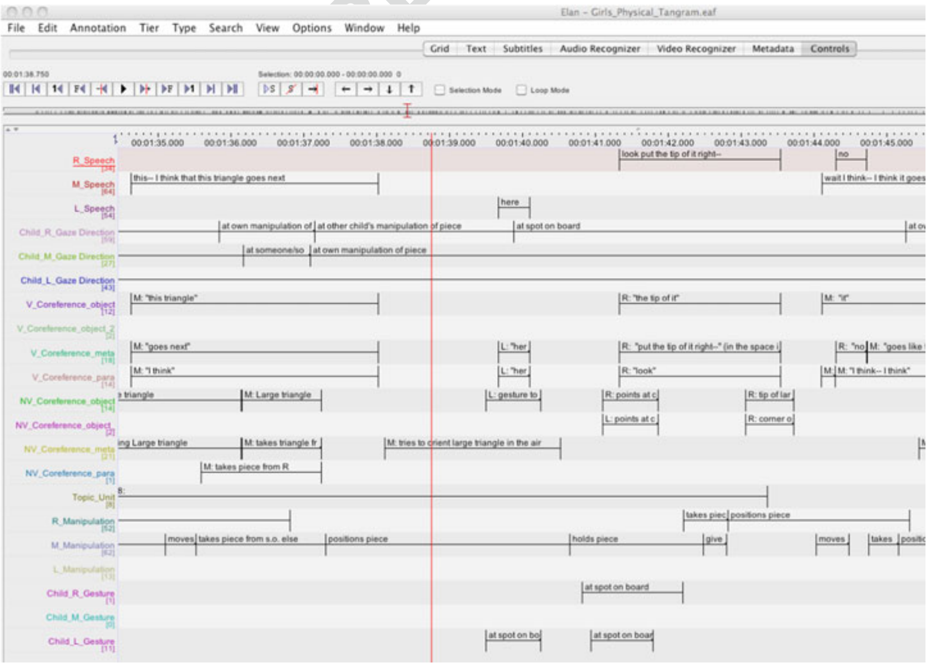


Fig. 5 Data from girls' physical setting as shown in Elan, allowing researchers to observe the synchrony of gestures, speech, and gaze

manipulation of each participant). We use gesture here to refer to pointing or indicating an object versus physically moving, or manipulating, it; this distinction helps to keep the live interaction embedded in the transcript. These three sets of tiers emerge directly from the video and ELAN allows us to display them with relative truth (within 0.1-0.5 s) to the moments at which they occur and overlap in the video recording—a capacity lost instantly in Excel, or linear, transcriptions. Transcribing from video to tiered annotations in ELAN allows us to keep the transcription dynamically close to the video itself, rather than isolating audio, visual, and temporal elements into separate elements in the process of translation to a linear transcript.

The next step is to analyze the verbal and nonverbal components of the scene, looking for cohesion to objects, subjects, and topics, as described in previous sections. The section of transcript featured in Figs. 4 and 5 captures the introduction of a new topic verbally and nonverbally by M. Looking vertically down the tiers, it is possible to read the moment as it unfolds, as if there are threads weaving through the component parts: M begins "this," referring to the large triangle R is holding and corrects herself to begin with "I think," establishing herself as a participant in the unfolding scene and then introducing the object and suggestion. Verbally, therefore, she is offering three potential points of cohesion for the group—the object of the triangle (which we analyze as an object-level reference), herself as participant (a para-level reference), and the concept of what goes next in the solving of the puzzle (a meta-level principle). Now, the nonverbal data available fleshes out this analysis considerably. Working from the transcript, we can observe that partway through M's statement she takes the triangle from R and R's gaze follows M's manipulation of the object (observe the shift in gaze on tiers 4–5 at 1:37.00). As M is manipulating the piece in midair (01:38–01:40.5, annotated in tiers 13, "nonverbal coreference—meta" and 17, "child M gesture"), which coheres to the meta-level chain of consideration about the piece's placement in the puzzle, L gestures to the board (transcribed on tier 12, "nonverbal coreference—object" and tier 21, "Child_L_gesture") and says "here" (at approximately 01:40, tier 3, "L_speech")—an ambiguous utterance, the significance of which could be easily lost or confused in a linear transcription of talk. A closer look, however, reveals that L's speech/gesture "here" doesn't demonstrate a particular location on the board as one might think from her choice of deictic, but rather functions as an attempt to gain access to the puzzle pieces and board. For the most part, M has been in control of the puzzle piece manipulation from the start of the exercise and L has had limited contact with the pieces. Her ambiguous gesture paired with her remark "here" indicates cohesion on the para level because, more than anything, it shows to M her interest in trying out *her* idea. A more detailed discussion of leadership negation follows this section.

We can read from the transcript that R's gaze follows L's gesture/utterance and that then, at 01:41, R joins the coalition by pointing directly to a spot on the board and uttering "look put the tip of it right—," which show coherence on object ("the tip" and "it" as well as gesturing to the place on the board), para ("look," drawing M's attention and echoing L's desire to participate), and meta (suggesting how the triangle should be oriented so that it will fit, cohering to the goal of fitting this triangle in within the lines and existing board structure) levels. At 1:41, when R joins in advising M, both R and L are gesturing to the same place on the board and R need not even finish her statement ("put it right—") as the content is already made clear by the shared focus of all three participants on object-, meta-, and para-levels. This largely unspoken cohesion around object, goal, and participants indicates the beginning of a coalition.

During 1:41–1:46, in merely 5 s, there is a sharply increased density of both nonverbal and verbal coreferences on object-, meta-, and para-levels by all three participants (see tiers

7–14). The annotations are briefer and bookend each other, indicating that while only one participant is actively speaking, they are all engaged in various modalities in communicating closely about what is occurring. A pivotal moment in the collaboration is initiated at 1:41.3 when R, in gesture and as mirrored in her speech, affirms L's suggestion and M's goal and takes the piece from L. The fact that the piece is freely given, as noted in the gesture tier, also testifies to this section as a period of active and participatory collaboration. At 1:48 in Fig. 5, the emptiness of the coreference tiers (7–14, the central tiers) indicate a slowed pace. There is a visible decrease in annotations of speech or gesture; M and R adjust the piece, but not much progress is being made; however, the richness of data available from the transcript allows us to reconstruct the story of this small moment. As the cohesion breaks apart—observe the shifts in L's gaze on tier 6 and the absence of coreferentially significant actions—the coalition that formed around shared points of discursive cohesion dissolves. The topic unit, which had focused the participants on the triangle, ends and there is a lull before the next topic unit is initiated, the next new piece or idea brought to the table. The metaphor of threads aptly describes the slow building of cohesion around multiple objects (the tangram piece itself and each place on the board is considered to be an object in space) and concepts that are interwoven and strung together across verbal and nonverbal dimensions of interaction, legible dynamically using an ELAN transcript.

Our system of multi-tiered ELAN transcription allows a determined reader to reconstruct a particular moment of cohesion and collaboration in a section of interaction. But, and perhaps more importantly, it allows us to identify the trends in the interaction that consistently coincide with the advancement of the problem-solving process. That is, we can actually "see" the peaks and flows of multi-level cohesion by scanning the coreference tiers. Zooming out to the macro level for a moment, the threads of continuity and sustained cohesion on the object, para, and meta-levels are observable on tiers 7–14. Reading across the tiers, tiers 7, 8, 11 and 12 show the utterances and gestures of L, M, and R (that is, all participants) cohering on the object-level (and in this particular section, these are all object-level references to the same shape, the large triangle, or to places on the board where that piece should go.) Between verbal and nonverbal annotations on the meta-level, actions of M and R are cohering on this level (tiers 9 and 13); and we have identified para-level cohesion between L and M (tiers 10 and 14). In this instance, the great number of coreferences on all three levels of discourse is indicative of distributed cognition that is driving the puzzle-solving forward: the participants are focused on the same conceptual problem, and are engaging each other directly. Thus, reading the multi-tiered ELAN transcript allows the researcher to make several observations: 1) trends in a single participant's involvement, for instance contributing primarily nonverbally or very frequently on a para-level; 2) periods of high-level collaboration where all participants are engaged in all three levels; 3) patterns of coreferential ordering in periods of high productivity, coalition formation, or the introduction of topic units, i.e. the trend of introducing topic units with para-level utterances/gestures that "sandwich" meta- and object-level cohesion; 4) patterns of involvement indicating dynamics of power, leadership, experience or confidence, i.e. the regular introduction of a new topic by a particular participant.

Now turn to the Excel transcript, which is derived from the ELAN transcription. While it is certainly easier to read, the synchronicity of action/utterances and richness of threads is lost; the Excel transcript evens out its contents. Nonetheless, where a multi-tiered ELAN document can be a struggle to read—in the screenshot captured in Fig. 5, many of the annotations are truncated for the reader and it is impossible to see both complete annotations and a sizeable chunk of transcript—the Excel transcript allows the speech/

JrnlID 11412_ArtID 9113_Proof# 1 - 09/03/2011

action annotations to be fully visible. Although the end of the topic unit was suggested in the previous ELAN example, it can be seen and annotated more clearly in the Excel transcript as beginning with the initiation of focus around a particular object (usually a tangram piece or another object like the mouse or seating arrangement) and ending with the last mention of that object/focus before the topic shifts to a new object/focus. The third column in the transcript, which captures transcribed coreferences, is no longer visually stacked as "threads" but is in this incarnation more useful as a countable or statistical measurement. Excel transcripts allow us to observe, across a number of topic units, a pattern of "sandwiching," that topic units often approximately begin and end with para-level coreferences and contain at least one meta-level coreference. (In the case of this section of transcript, M begins and then self-corrects to begin her statement with "I think"—a para-level self-reference.)

In addition to highlighting the ordering of coreferences and improving the legibility of speech and actions, Excel transcripts allow us the opportunity and space to give a short narrative of each topic unit. The description shown in the "Focus" column of Fig. 4 helps to keep the analysis embedded in the unfolding sequence of events, rather than allowing analysis to become increasingly removed from the intuitively observable arc of the interaction. Unlike the "Speech/Action," "Gaze," and "Medium of cohesion" columns, we filled out the "Focus" column by returning to the video and describing directly from there per each section determined as a "topic unit" in the ELAN analysis. We determine "focus" based on where the gaze, speech, and action of each participant is oriented; so while some participants may sustain focus on the same task for a long period of time, others, particularly in the physical setting, shift their focus from object to object more rapidly. Where all three participants are sharing focus (i.e. involved in a coalition) the focus column is three columns wide; in other sections of the interaction where focus is divided, the "focus" column may be divided into two or three different columns, showing smaller coalitions and periods where each individual is focused on a separate task. (For a visual example of divided focus represented in an Excel transcript, see Fig. 6.) This column is extremely valuable in not only identifying coalitions, but also helping researchers to understand what conditions motivate the formation and breakdown of coalitions and also to note particular patterns in each individual's participation. Furthermore, the focus column mitigates the complexity and dynamism of the interaction by providing a visual representation of the participants' shifting focus.

The "common focus" column serves to synthesize data gathered in the "focus" column, indicating numerically how many participants are sharing focus or acting as a coalition on a task at a given time. This column is unique to our Excel transcript and has proven very valuable in comparing male and female settings and virtual vs. physical settings, as the frequency and duration of three person coalitions varies widely. The last two columns, "Topic initiator" and "Coalition initiator" are also both unique to our Excel transcript and immensely useful in giving us data on dynamics of leadership and trends in the dynamics of the group. "Topic initiator" allows us to note which participant initiated the topic unit (usually by introducing an object or topic such as the rules of the game, seating arrangement, or whose turn it is with the mouse; in the example from Figs. 4 and 5, M initiates the coalition by introducing the large triangle as the one that "goes next"). The "Coalition initiator" column indicates which participant picks up the topic introduced by cohering on at least the object level (and usually para- or meta-level as well); or, in cases such as our example where both participants join the coalition of shared focus on the triangle, the order in which they join. (In our example, L initiates a coalition by cohering verbally and nonverbally to the topic of the large triangle by saying "here" and pointing to a

Focus (S/J/A) Student S	Focus (S/J/A) Student J	Focus (S/J/A) Student A	Topic Unit	Number of participants with common focus	Initiator of TU	Initator of coalition
S and J are opening bag	S and J are opening bag	A focuses on the sheet of paper	#1 (S,J,A) Opening bag	2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2	J	S
A sets large triangle down in center, and S responds by implying that the second large triangle won't fit. The problem is resolved when A slides both triangles off the outline. J watches.			#2 (S,J,A) Large triangles	3, 3, 3, 3, 3, 3, 3, 3, 3, 3	A	S
J pulls a small tri. out of the bag; he tries to place it but then withdraws and keeps it in his hand	A focuses on placing parallelogram and S watches and encourages.	A focuses on placing parallelogram and S watches and encourages.	#3 (A,S) Parallelogra m	3, 2, 2, 2, 2, 2	A	S
J pulls a small tri. out of the bag; he tries to place it but then withdraws and keeps it in his hand	S hands square to A and then watches him place it.	S hands square to A and then watches him place it.	#4 (S,A) Square	2, 2, 2, 2, 2	S	A
S puts large triangle in center, J watches	S puts large triangle in center, J watches	A focuses on other	#5 (S,J,A) Large Triangle	2, 2		
A, J, and S adjust pieces to eliminate white space. There is some disagreement about whether it fits but then they agree that it does.			#5 (S,J,A) Large Triangle	3, 3, 3, 3, 3, 3	S	J

Fig. 6 Physical data for boys presented in Excel shows clear topic units and formations of coalitions, with shared focus between 2–3 participants at all times. "Focus (S/J/A)" columns allow narrative of discourse to emerge and visually represent cohesion of focus. (50 s)

place on the board. These actions refer, or more accurately "corefer," to the task introduced by M.) Based on these two columns, we are able to recognize patterns in an individual's behavior as well as *quantify* his or her contributions to the group. We can also see how leadership roles develop and crystallize over the course of the puzzle-solving process. The

"Focus," "Common Focus," "Topic initiator," and "Coalition initiator" columns are made possible by the layout of Excel, in which they synthesize information from the analysis conducted in ELAN and make it easy to scan and look for patterns. The utility of these columns motivates us to continue to produce Excel transcripts following our ELAN transcripts.

Challenges and limitations of coreference-based multimodal coding

A number of ambiguities emerged as we developed and worked with this coding system, which is to be expected given that the coding reflects intuitively felt dynamics of collaboration, yet its strength is in giving researchers the analytical tools to break down moments closely and observe dynamics, coalitions, and patterns as they form as part of cooperative knowledge construction.

Because the temporal scale of our Excel transcript is determined by the number of annotations (speech, gesture, or manipulation) added to the ELAN tiers, time is somewhat skewed in the Excel document relative to how much recordable activity occurs within that period. This can be both useful and misleading. One gets the sense from the Excel transcript that activity is constant because actions are added to each line regardless of the time at which they occurred. The ELAN transcript, on the other hand, allows the researcher to see "empty" space (i.e. periods of little or no interaction), providing a more accurate depiction of the fluctuating degree of cooperation. The appearance in the Excel transcript of increased relative activity between 1:39 and 1:41 corresponds with our prior observation that between 1:41 and 1:46 there is a pivotal peak in discursive cohesion, but it doesn't stand out, whereas in ELAN we can see a cluster of annotations preceded and proceeded by blank space. In Excel, however, researchers can easily spot recurring patterns in coreferential behavior because the sequence of actions is stressed more than anything else. The tension between readability and accuracy in the transcripts is ongoing.

We also encountered difficulties during the coding process itself. Recall that we specified three distinct levels of coreferences: object, meta and para. In fact, we discovered many instances where an utterance or gesture could function on multiple levels of discourse. Object-meta (e.g. [R adjusts two triangles to fit together inside the lines] or "that (*obj*) doesn't fit (*meta*)") and object-para coreferences (e.g., [M takes parallelogram from L] or "hey, look (*para*) here (*obj*)") contained within the same gesture or utterance are quite common. The overlapping and multifunctionality of verbal and nonverbal utterances supports our theory that coreferences build on each other, forming chains of cohesion that comprise topics in the discourse. For instance, in the previous section ("A more detailed example of multimodal coding") we discussed the coding of an ambiguous utterance and gesture and the need for context to understand. Another example can be found in a moment where a student suggests switching two triangles and another student says, "They're the same." While at first we read this as an object-level coreference, referring again to the two triangles as points of cohesion, we observed that her utterance functions in the discourse as a directive *not* to switch the two pieces, based on an observation of their geometric identity. Therefore, the utterance functions as both an object-level coreference and a mathematic meta-level coreference in the discourse.

Phrases like "I think" and "I think we should" occur so frequently that we were at first hesitant to identify them as containing self-referential deixes, but in many cases such utterances did function to introduce the self-referring individual into a coalition of focus where he/she and his/her thoughts became established as a point of cohesion for future discourse. Additionally, we encountered ambiguity around questions, such as "it looks like a ship, doesn't it?" and commands, such as "go back," "look," and "wait." Questions seem to take both the self and the asked as implicit points of potential cohesion, just as commands seem to operate on a para-level (referring to the commanded person and their actions) as well as meta-level (orchestrating the sequence of activity in solving the puzzle.) We ultimately decided to code these based on function in context—that is, when a question or command was cohered to by the addressed participant or the group, we coded it as a para-level reference.

In a sense, it is easy to lose the forest for the trees when coding coreferences. Focusing in on a particular deictic often skews one's perception of its function in the discourse as a whole. We challenged ourselves to flesh out the function(s) of a single utterance or gesture and were almost always able to arrive at consensus among coders. In a blind comparison, 84% of all our annotations agreed, and that figure rose to 96% after discussion. We were able to keep the rigor of our coding at a reliable level by coding separately and discussing ambiguities and, when coding alone, by referring frequently to the video both when transcribing to ELAN and also when transcribing to Excel. This process of double-checking (that is, rechecking our analysis of discursive cohesion during the transcription process to Excel) challenged us to return many times to the video in order to arrive at a coreferential analysis that is as accurate as possible.

Another limitation of using coreferences to describe group problem solving behavior is that we can only code references that are *shared*. There were often times when a participant had a puzzle piece in her hand and experimented by flipping it or rotating it *without* placing it on the board or showing the piece to the others. While this activity is important to the individual participant's understanding of geometric properties, it isn't part of the shared problem solving process and thus doesn't fit into the coding scheme. Alternatively, one virtuoso problem solver may advance the progress of the puzzle significantly, but without communication (gestural or verbal) with the other participants, we would annotate his/her actions and utterances recognizing the lack of coreference (or distributed cognition) occurring. A brief utterance, such as a "yes" or "no" or "{gasp}" might be part of a larger chain of discourse that coheres to a particular topic; on the other hand, there are utterances and gestures that are *not* part of a larger chain of discourse, do not influence another participant, and so are not coded as coreferences. Adhering to this commitment to draw boundaries between distributed and individual cognition proved particularly challenging when we were coding the virtual puzzle setting, since the group members' gazes were almost always focused on the shared screen, but shared gaze does not necessarily indicate that distributed cognition is occurring. Identifying coalitions in the virtual setting was therefore more difficult, and would be aided by the integration of more advanced gaze-tracking software.

Comparatively, topic units and coalitions are much easier to identify and trace because they arise from the analysis of coreferences and because their periodic function is intuitively visible. The "focus" column of our Excel transcriptions helped us to maintain rigor in identifying the initiation and disintegration of both coalitions and topic units. Initially, we identified topic units as periods of focus on a single object-oriented task (usually the placement of a particular piece) and then recognized topic units that occurred around meta-level principles, such as adjusting the puzzle to fit within the lines, how to rotate a piece on the computer, or whose turn it is to use the mouse on the virtual puzzle. We also recognized that there are periods where there are no topic units occurring; that is, where there is not a coalition of shared focus around any single task.

Additionally, we were challenged to bring the apt visual metaphor of discursive threads to our coding. We struggled to indicate in ELAN and Excel when a coreference was linked to the same object, as opposed to proximate coreferences to different objects. We considered using a system to code coreferences to particular shapes by annotating object-level coreferences with a number (e.g. "1" to represent the large triangle, etc.) but it became too unwieldy. Our coding system could easily accommodate such an implementation of specifications of coreferences and we see much potential for such an implementation to be used to calculate statistics and produce visuals of threads of cohesion.

In this study, we are limited from making broad generalizations about patterns of interaction or the influence of gender, or TEMA competency by the small cross-section of data we transcribed (four four-minute sections). However, this amount of data was sufficient for us to apply and tweak our method of transcription and representation to be reliable and to testify to its potential for identifying patterns of interaction in CSCL environments. We also identified room for improvement in our study through the use of a system of gaze tracking for students working with a tangram puzzle on a computer. We "transcribed" gaze from the videos, distinguishing between moments when the child was looking at another person, the computer/table, or off the camera frame (perhaps at the researcher or something else in the room). Advanced gaze tracking software and hardware is available which could reflect where the child is looking on the computer screen, i.e. at which shape. This indication of focus would be an immense addition to our understanding of coalitions and coreferences, since we treat gaze as a primary indicator of focus.

Initial findings of our research

Perhaps the most powerful finding of our research was the possibility of using widely available software to apply the theoretical framework of coreferences to reliably and rigorously trace distributed cognition and discursive cohesion to objects, participants, and geometric and cooperative concepts. This method of analysis can be applied to a wide range of collaborative situations and has potential for revealing patterns of discursive cohesion in interactions that are oriented to problem solving, whether computer-oriented or not. In this section we will discuss some of the emergent patterns we have observed in our data.

As stated above, for both the physical and virtual settings, we have observed a concentration of meta and para coreferences at the beginning and ending of topic units (that is, heightened discourse about the puzzle and the participants' relation to the puzzle), suggesting that the children are sensitive to a step-by-step approach to problem solving. Establishment of coreferences to mathematical principles, the wider context of the new topic in the larger context of puzzle-solving, and the individual participants contributing mark the initiation of a coalition of focus and distributed cognition necessary for successful problem-solving. We found that in the CSCL/virtual setting, there was a greater number of para and meta coreferences and more frequent coalitions of focus among all three participants. Figures 6 and 7, data from the boys' physical and boys' virtual settings from the first 50 s of activity, display some of these trends. In the physical setting (Fig. 6), there is less consistent focus but the participants are all engaged in initiating coalitions and topic units. In the virtual setting (Fig. 7), there is consistently more shared focus among all three participants but the initiation of topic units and coalitions tends to be less balanced. We found that students who frequently initiated topic units and coalitions, intuitively observable as being the most active or dominant figures, tend to be *more* active/dominant in the virtual setting than the physical setting. (In the girls' setting, one participant initiates 45% of the topic units and 38% of the coalitions in the virtual setting as compared to 28% of topic units and 0% of coalitions in the physical setting; this is also reflected in the preponderance of participation by L and R as initiators in the virtual setting, as shown in

Topic Units	other TU's (non-puzzle)	shared focus	Initiator of topic unit	Iniatiator of coalition
		0		
		0	M	(no coalition)
	#1: (L) L takes mouse, R reaches for mouse.	2	L	R
		2		
		2		
		3		
#2: Square: L (mouse) picks up square, initiating topic. R suggests putting square in the middle and L says that the square is the head. M watches.		3	L	R
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
		3		
	attempt to take mouse (ignored)	2	R	(no coalition)
Large Triangle: L picks up large triangle but then discards it without comment from M or R. (ignored)		2	L	(no coalition)
		2		
	RS asks R to sit down.	2		
		2		
#3: Medium Triangle: R introduces topic (medium tri.), L suggests location for the piece and places it in the hip region. M watches.		3	R	L
		3		
		3		
		3		
	L suggests puzzle is a guy kicking. (ignored)	3		
Small Triangle: L initiates, places small triangle in right leg, M&R watch (ignored)		3	L	(no coalition)
		3		
		3		

Fig. 7 Virtual data for boys presented in Excel shows more disjointed topic units, more frequent "failed"/ ignored introductions of new topics, and fewer coalitions overall but more consistently shared focus among all 3 participants compared to same time period (50 s) with same group of participants as in Figure P

Fig. 7, and the virtual disappearance of participant M). Less "dominant" figures, i.e. students who are *least* active in initiating topic units and coalitions, who are often not part of a coalition of focus, and/or whose initiations of topic units are ignored by the other students, tend to participate more in the physical settings than in the virtual settings. (For instance, one female student initiated 50% of coalitions in the virtual setting and 86% in the physical setting; a male student initiated 11% of coalitions and 0% of topic units in the virtual setting as compared to 21% of coalitions and 21% of topic units in the physical setting.) This trend is intuitively observable by teachers as well as being observable through analysis of topic units, focus, and coalition formation. Coreferential and coalition analysis foregrounds group dynamics and dynamics of leadership by grounding leadership in the establishment and use of coreferences.

We are especially interested in the variation among the four data sets in meta coreferences to (1) geometric or mathematical principles (and the properties of the puzzle pieces), and (2) collaborative problem-solving strategies. Our initial observations suggest that, for the group of female students, there are proportionally more coreferences of the collaborative strategy type in the virtual setting than in the physical setting. This may be explained in part by the fact that the participants had to negotiate control over the mouse, and were thus more aware of the parameters of cooperative group activity. The female students are highly attuned to the discourse of turn taking in the virtual setting; each student gets a turn with the mouse and the discourse reveals that they have a sense of how long a turn should be. Turn taking is brought up only very briefly in the boys' virtual setting and one male student has control of the mouse throughout the puzzle. While we are hesitant to attribute the appearance of patterned dynamics directly to gender differences, tested math competency, or to previous experience with tangram puzzles or classroom group problem-solving experience, the data suggests that although there are consistent trends in the difference between virtual and physical manipulatives, these other variables also affect student interaction in fairly consistent, predictable ways.

A comparison to extant multimodal approaches in CSCL

Cakir et al. (2008) also focus on the mechanics of cooperation and explore the content of Q13 coreferentiality through their study of how math students working in synchronous chat environments to solve math problems "achieve intersubjectivity and shared cognitive accomplishments" and group organization (p.3). Their work reinforces the concept of the necessity of joint problem solving space as the foundation of group cognition and the usefulness of combined ethnographic methods and interactional/discourse analysis as a way to understand interwoven references.

One finding of their work is the centrality of the visual realm (the whiteboard is the "dual space" on which students are working and seeing others' work) as the source of references and primary interactional resource, particularly in the absence of gesture as a medium for creating and mirroring transient images in space. The sequential nature of problem-solving and the continual formation of historical context (or, the laying of an indexical field for coreferentiality) for joint work translates from the chat, purely virtual setting to the physical and computer-mediated settings that we describe here (p.22). The construction of historical content is also addressed by the work with interactive whiteboards in classroom settings (Kershner et al. 2010), who used interactional discourse analysis to confirm that turn-taking and role-switching, including periods of silent listening and watching, are indeed taken forward into subsequent interactions. We can offer to the observations made by this group the methods we have developed for more fine-grained and temporally sensitive analysis of gaze and gesture.

Conclusion and implications for future research

To sum up, the focus of our inquiry was on how we could identify and trace distributed cognition through discursive cohesion on multiple levels in CSCL and non-computer-supported collaborative learning environments. We examined the moment-to-moment details of speech, gesture, object manipulation, and gaze as ways that participants construct a shared indexical ground for future interactions. We analyzed these details of interaction as coreferences to objects, places, prior content of discourse, individuals, the group, and mathematical and collaborative principles; and we explored how chains of coreferences form topic units and coalitions of focus that are necessary for collaborative problem-solving and distributed cognition. We also created transcripts of our data that rigorously retain substantial information about different dimensions and allowing the reader to observe the moment-to-moment dynamism of overlapping coreferential chains. Our observations about distributed cognition and how instructional technology (manipulatives) mediates childhood learning and collaborative patterns open many doors for future work. As Barron (2003) notes, by focusing on the group CSCL researchers are able to describe "interactions that capture the dynamic interplay in meaning making over time in discourse between participants, what they understand, the material resources they use…and how they are taken up or not in a given discourse" (p. 6). Our work aligns with this sentiment as well as with a latter one that states that our findings should, ultimately, be translated so that teachers can diagnose and support collaborative learning (p.48).

Thus, our initial position, drawn from empirical work on mathematical reasoning, proposed that tangram puzzle activities (irrespective of physical or virtual qualities) provided opportunities to create and share meaningful artifacts socially as learners could easily be organized into small groups for the duration of an instructional exercise. Once learners are organized as such, we might entertain the application of Vygotsky's Zone of Proximal Development (ZPD) as described by Chaiklin (2003). Vygotsky (1978) proposed that higher order psychological functions are produced first in social interaction before being internalized by individual students, supporting the significance of coordinating individual cognitive processes with the social processes of a community to produce individual learning. Based on previous efforts, we found that Vygotsky's theory served as a bridge between traditional, individual learning theories and constructivist, social learning theories; each student has his or her ZPD and can only function within this zone with the assistance of a more experienced learner(s). These particular social configurations shape how students learn mathematical reasoning, as explained by Enyedy:

> By aligning one's individual participation with the ongoing organization of a distributed system that extends beyond the individual's mind, that individual eventually learns how to perform these same functions competently when other aspects of the system are absent (Cazden 1997; Vygotsky 1978; Wertsch 1985; Wertsch and Stone 1999). The means for development, then, is sustained social interaction and the continual shift toward taking more responsibility for one's own activity (2003, p. 364).

We sought to create a model of analysis for the "ongoing organization" of distributed cognition and to use our case study to begin to answer the question, what are the markers and patterns of organization of distributed cognition in a collaborative learning setting? In our case study of two groups of 7- and 8-year old students solving tangram puzzles in tabletop and computer-supported settings, three students worked to solve a tangram puzzle, communicating about the geometric properties of shapes, the rules of the puzzle, their relationships to each other, and their progress. As students worked to complete tangram puzzles, often a group "leader" emerged, and peers looked to this more advanced individual for guidance. The ability to learn through imitation and adaptation of the leader's actions allowed students the opportunity to formulate and refine psychological functions that, when working independently may only be in the earliest stage of development and therefore cannot be effectively performed (Chaiklin 2003). The successful solutions of the tangram puzzles by students in our study were not the mental achievements of a single individual but a group accomplishment carried out through the coordination of shared meanings and coreferences to both objects and mathematical and collaborative principles. The deeply collaborative nature of problem-solving is demonstrated and testified to by our system of analysis which foregrounds shared reference points as coreferences, shared periods of focus, and the formation of coalitions.

Furthermore, our case study demonstrated that small group collaboration and the creation/sharing of artifacts operates differently in the virtual and physical realms; learners are more likely to discuss and articulate their ideas in the computer-supported setting, but there is a decrease in gestural communication and so students who are more hands-on learners may fade into silence, as happens in both the boys' and girls' computer-supported settings. While we had predicted the natural emergence of group leaders, our observation that they tend to be more dominant in the computer-supported setting brings us to question the simplicity of Chaiklin's (2003) argument about learning through imitation and adaptation. While the computer-supported setting increases higher-level communication about mathematical principles and the properties of shapes, the physical tangram-solving setting offers increased opportunities for mirroring gestural imagery and adopting habits of manipulating pieces to contribute to the whole. Our case study does not predict that these patterns will be played out in all mathematical computer-supported learning contexts, but rather demonstrates the potential of coreferential multimodal coding for deriving information about how distributed cognition is accomplished in a CSCL setting. In this paper, we have demonstrated the efficacy of using a system of multimodal coreferential coding for tracking and measuring distributed cognition and its relation to interactional dynamics and dominance.

References

Ares, N., Stroup, W., & Schademan, A. (2009). The power of mediating artifacts in group-level development of mathematical discourse. *Cognition and Instruction, 27*(1), 1–24.

Arzarello, F., Paola, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Mathematics, 70*, 97–109.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences, 9*, 403–436.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.

Bjuland, R., Cestari, M., & Borgersen, H. (2008). The interplay between gesture and discourse as mediating devices in collaborative mathematical reasoning: A multimedia approach. *Mathematical Thinking and Learning, 10*(3), 271–292.

Bohning, G., & Aithouse, J. K. (1997). Using tangrams to teach geometry to young children. *Early Childhood Education Journal, 24*(4), 239–242.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (2nd ed.). Washington, D.C.: National Academy.

Brugman, H. & Russell, A. (2004). Annotating multi-media / multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation.

Cakir, M. P., Zemel, A., & Stahl, G. (2010). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 115–149.

Cannon, L., Dorward, J., Duffin, J., Heal, B., & Foletta, G. (2004). National library of virtual manipulatives. *Mathematics Teacher, 97*(2), 158–159.

Cassell, J., & McNeill, D. (2001). Gesture and the poetics of prose. *Poetics Today, 12*, 375–404.

Cazden, C. B. (1997). Performance before competence: Assistance to child discourse in the zone of proximal development. In M. Cole, Y. Engestrom, & O. Vasquez (Eds.), *Mind, culture, and activity: Seminal papers from the Laboratory of Comparative Human Cognition* (pp. 303–310). New York: Cambridge University Press.

Chaiklin, S. (2003). The zone of proximal development in vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), *Vygotsky's educational theory and practice in cultural context* (pp. 39–63). Cambridge: Cambridge University Press.

Clements, D. H., Sarama, J., & DiBiase, A.-M. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah: Lawrence Erlbaum.

Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research, 13*(1), 9–19.

Dietz, P. & Leigh, D. (2001). *DiamondTouch: A multi-user touch technology.* In Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (Orlando, Florida, November 11–14). UIST '01. ACM Press: New York, NY, pp 219–226.

Enyedy, N. (2003). Knowledge construction and collective practice: At the intersection of learning, talk, and social organizations in a computer-mediated mathematics classroom. *The Journal of the Learning Sciences, 12*(3), 361–407.

Gehlen, A. (1988). *Man, his nature and place in the world*. New York: Columbia University Press.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Austin: Pro-Ed.

Glover, D., Miller, D., Averis, D., & Door, V. (2005). The interactive whiteboard: a literature survey. *Technology, Pedagogy and Education, 14*(2), 155–170.

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics, 49* (2), 193–223.

Granott, N. (1993). Patterns of interaction in the co-construction of knowledge: Separate minds, joint effort, and weird creatures. In R. H. Wozniak & K. W. Fischer (Eds.), *Development in context: Acting and thinking in specific environment* (pp. 183–207). Hillsdale: Lawrence Erlbaum Associates.

Han, J. Y. (2006). *Multi-touch interaction wall*. In SIGGRAPH '06: ACM SIGGRAPH 2006 Emerging technologies. ACM Press: New York, NY, 25.

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., & Rogers, Y. (2009). *Around the table: Are multiple–touch surfaces better than single–touch for children's collaborative interactions?* In Proceedings of CSCL '09. Mahwah, NJ: Lawrence Erlbaum Associates.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(2), 174–196.

Husserl, E. (1931). *Ideas: General introduction to pure phenomenology (W. R. B. Gibson, Trans. Third Edition, 1958)*. London: Allen & Unwin.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge: MIT.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265–288.

Hutchins, E., & Klausen, T. (1998). Distributed cognition in an airline cockpit. In Y. Engestrom & D. Middleton (Eds.), *Cognition and communication at work* (pp. 15–34). New York: Cambridge University Press.

Kafai, Y. B. (2006). Constructionism. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 35–46). Cambridge: Cambridge University Press.

Kendon, A. (2008). Some reflections of the relationship between 'gesture' and 'sign'. *Gesture, 8*, 348–366.

Kershner, R., Mercer, N., Warwick, P., & Staarman, J. K. (2010). Can the interactive whiteboard support young children's collaborative communication and thinking in classroom science activities?. *International Journal of Computer-Supported Collaborative Learning*, *5*(4).

Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge: Cambridge University Press.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity, 7*(4), 273–290.

McNeill, D. (2006). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, D. (2009a). Gesture as a window onto mind and brain, and the relationship to linguistic relativity and ontogenesis. In C. Müller, E. Fricke, et al. (Eds.), *Handbook on body, language, and communication*. New York: Mouton de Gruyter.

McNeill, D. (2009b). Thinking while speaking and gesturing: Language and gesture as a dynamic and integrated system. In C. Müller, E. Fricke, et al. (Eds.), *Handbook on body, language, and communication*. New York: Mouton de Gruyter.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking for speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: CUP.

McNeill, D., et al. (2010). Mind Merging. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, language, cognition, and identity* (pp. 143–164). London: Taylor and Francis.

Merleau-Ponty, M. (1945). *Phénomenologie de la perception*. Paris: Gallimard.

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics, 47*(2), 175–197.

Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathmatics, 8*, 372–377.

Moyer-Packenham, P. S., Salkind, G., & Bolyard, J. J. (2008). Virtual manipulatives used by K-8 teachers for mathematics instruction: Considering mathematical, cognitive, and pedagogical fidelity. *Contemporary Issues in Technology and Teacher Education, 8*(3), 202–218.

Noss, R., & Hoyles, C. (2006). Exploring mathematics through construction and collaboration. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. New York: Cambridge University Press.

Papert, S. (1980). *Mindstorms*. New York: Basic Books.

Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. New York: BasicBooks.

Pea, R. (1993). Practices of distributed intelligence and design for education. In G. Salomon (Ed.), *Distributed cognition*. Cambridge: Cambridge University Press.

Radford, L., Edwards, L., & Arzarello, F. (2009). Introduction: Beyond words. *Education Studies in Mathematics, 70*, 91–95.

Rekimoto, J. (2002). *SmartSkin: An infrastructure for freehand manipulation on interactive surfaces*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '02 ACM Press, New York, NY, pp 113–120.

Rick, J., & Rogers, Y. (2008). *From DigiQuit to DigiTile: Adapting educational technology to multi-touch table* Paper presented at the IEEE Tabletops 2008 and Interactive Surfaces 2008.

Romeo, G., Edwards, S., McNamara, S., Walker, I., & Ziguras, C. (2003). Touching the screen: Issues related to the use of touchscreen technology in early childhood education. *British Journal of Educational Technology, 34*(3), 329–339.

Roth, W.-M. (2001). Gestures: Their role in teaching and learning. *Review of Educational Research, 71*(3), 365–392. doi:10.3102/00346543071003365.

Salomon, G. (Ed.). (1993). *Distributed cognitions: Psychological and educational considerations*. New York: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265–283.

Shen, C., Vernier, F. D., Forlines, C., & Ringel, M. (2004). DiamondSpin: An extensible toolkit for around-the-table interaction. Proc. of CHI'04, 167–174.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge: MIT.

Strijbos, J. W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small group chat communication. *Learning & Instruction. Special issue on measurement challenges in collaborative learning research, 17*(4), 394–404.

Tapper, J. (2007). Tangrams and geometry concepts. *Connect Magazine, 20*(4), 7–11.

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 229–258). Mahwah: Lawrence Erlbaum Associates, Inc.

Van de Walle, J. A. (2001). *Elementary and middle school mathematics: Teaching developmentally* (4th ed.). New York: Longman.

van Hiele, P. M. (1999). Developing geometric thinking through activities that begin with play. *Teaching Children Mathematics, 5*(6), 310–316.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge: Harvard University Press.

Wertsch, J. V., & Stone, C. A. (1999). The concept of internalization in Vygotsky's account of the genesis of higher mental functions. In P. Lloyd & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments* (Vol. 1, pp. 363–380). Florence: Taylor & Francis/Routledge.