Summer 8-17-2019

The Effect of Peer Collaboration on Kinematic Problem Solving

Alyssa DeRonda
derondal@shu.edu

Recommended Citation

DeRonda, Alyssa, "The Effect of Peer Collaboration on Kinematic Problem Solving" (2019). *Seton Hall University Dissertations and Theses (ETDs)*. 2693.
https://scholarship.shu.edu/dissertations/2693

THE EFFECT OF PEER COLLABORATION ON KINEMATIC PROBLEM SOLVING

by

Alyssa DeRonda

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Master of Science in Experimental Psychology
with a Concentration in Behavioral Sciences

In

The Department of Psychology
Seton Hall University
August, 2019

SETON HALL UNIVERSITY

College of Arts & Sciences

APPROVAL FOR SUCCESSFUL DEFENSE

Masters Candidate, Alyssa DeRonda, has successfully defended and made the required modifications to the text of the master's thesis for the M.S. during this Summer Semester 2019.

THESIS COMMITTEE

(please sign and date beside your name)

Mentor:
Amy Joh, Ph.D.:____

Committee Member:
Kelly Geodert, Ph.D.:____

Committee Member:
Fanli Jia, Ph.D.:____

Acknowledgements

I would like to extend my sincerest gratitude to my thesis advisor, Dr. Amy Joh. Without her support, patience, and encouragement I would not have been able to successfully complete this research thesis. She taught me to think critically and challenged me to reach my fullest potential.

I would like to thank my research team, Hailey Holt, Catherine Rubens, and Nina Danner, for all their hard work. I would not have been able to complete my study without their help.

I would like to thank all my friends who have provided words of encouragement throughout this difficult process. A special thank you to Jessica Perez, Kimberly Poss, Kathrina Alapatt, Adrianna Hughes, Nicholas and Katie Malin, and my roommate, Sean Bogart, who guided me through this adventure and many others.

Lastly, I would like to thank my family for giving me the many opportunities that led me to pursue what I love.

Table of Contents

List of Figures

List of Tables

Abstract

Kinematic problems, a type of physics problem that involves object motion, pose a challenge for adults (Caramazza, McCloskey, & Green, 1981; Kozhevnikov, Motes, & Hegarty, 2007; McCloskey, 1983b; McCloskey, Washburn, & Felch, 1983). Adults often incorrectly predict the path of a moving object despite having prior experience with moving objects or formal physics education (Caramazza et al., 1981; Kaiser, Jonides, & Alexander, 1986). One way to improve kinematic problem solving may be through peer collaboration. Working together with a partner to solve a problem allows both people to help each other remember important parts of a complex problem and discuss different perspectives (Dimant & Bearison, 1991; Fawcett & Garton, 2005; Kozhevnikov & Thornton, 2006; Vygotsky, 1978). The current study investigated whether peer collaboration can improve kinematic problem solving by evaluating adults' performance on near and far transfer tasks after completing kinematic practice problems. Of special interest was the use of spatially-oriented language. Participants were assigned to one of three practice conditions: Collaborative, Alone-Talk, or Alone-Quiet. Results showed that peer collaboration did not affect performance on practice problems or near and far transfer tasks. However, analysis of spatially-oriented language revealed that specific types of language were positively correlated with accuracy on the near transfer task.

Keywords: kinematic, peer collaboration, near transfer, far transfer

The Effect of Peer Collaboration on Kinematic Problem Solving

Daily interactions with moving objects are kinematic problems, a type of physics problem involving predicting the path of a moving object (Kozhevnikov, Hegarty, & Mayer, 2002; Kozhevnikov et al., 1999). Many daily kinematic tasks, like driving a car or walking through a crowded street, present novel situations related to tracking an object's motion, anticipating changes in its direction or speed, and making predictions about its trajectory. Surprisingly, despite extensive everyday experience, we make erroneous predictions about moving objects. For example, when asked to draw the trajectory of a ball as it rolls down a ramp and off a cliff, adults provide a number of different answers including responses that defy physical laws (McCloskey et al., 1983). The most frequent error involves dismissing the horizontal movement of the ball and assuming the ball will fall straight down once it rolls over the cliff. Another incorrect response involves predicting that the ball will continue horizontally and then fall straight down, creating an inverse L-shaped path—like how Wile E. Coyote fell off cliffs in cartoons. Similar results have been found for other kinematic problems such as a ball being dropped from a plane or conveyor belt, or a hockey puck being kicked as it drifts across the ice (Caramazza, McCloskey, & Green, 1981; Champagne, Klopfer, & Anderson, 1980; Clement, 1982; Kozhevnikov, Motes, & Hegarty, 2007; McCloskey, Washburn, & Felch, 1983).

Poor kinematic reasoning can have serious consequences. Luckily, there is evidence to suggest that kinematic reasoning may be improved by improvements in spatial visualization, the ability to mentally represent spatial relations (Kozhevnikov, Motes, & Hegarty, 2007). A key to improving spatial visualization is learning to use mental imagery to track movements while avoiding a concurrent increase in cognitive load (Isaak & Just, 1995). One way to achieve this feat may be through peer collaboration. There is evidence that suggests working with a peer improves critical thinking when completing a spatial task (Phelps & Damon, 1989), perhaps through the internalization of new perspectives (Hogan & Tudge, 1999; Tudge, 1992; Vygotsky, 1978; Vygotsky, 1978). These are skills that can improve the representation of spatial relations, and working with another person means each person can take charge of a component in a problem to reduce cognitive load. Therefore, the purpose of this thesis was to investigate the potential facilitative effects of peer collaboration on kinematic problem solving.

Kinematic Problems and Common Errors

In studies with kinematic problems, participants are typically presented with images such as a person dropping a ball while walking on flat ground or water flowing out a curved tube. The participants' task is to extrapolate an object's trajectory based on its current path and other relevant information (e.g., Caramazza, McCloskey, & Green, 1981; McCloskey, 1983a; McCloskey, Caramazza, & Green, 1980; McCloskey et al., 1983). Researchers have used kinematic problems because formal knowledge of physics is not required to solve these problems. Additionally, responses can be evaluated quantitatively and qualitatively from the direction and shape of the predicted path.

A variety of kinematic problems are used to investigate kinematic reasoning and problem solving. The most straightforward problems are rectilinear problems that include one vector of motion (movement along a straight line) such as rolling a ball in a straight line or tossing a ball straight up in the air from a fixed point (Clement, 1982; Kozhevnikov & Hegarty, 2001). Rectilinear problems can reveal beliefs and reasoning about object motion and force but not path because most respondents can accurately predict the trajectory of an object with one vector of motion.

Instead, projectile motion problems (movement along a parabolic curve) are favored by researchers because they require reasoning about two vectors of motion, horizontal and vertical, to accurately predict the trajectory of an object. Participants are typically asked about a vehicle transporting an object in a specified direction then releasing that object at a given position such as a swinging pendulum made of a ball and string, a walking person holding a ball, a ball attached to a conveyor belt traveling over a canyon, and a ball rolling down a ramp positioned at the edge of a cliff (Caramazza et al., 1981; McCloskey et al., 1983). In each case, participants must consider the object's existing horizontal motion and incorporate gravity to predict the object's change in direction. Other problems depict a moving object experiencing an external force from another direction such as a sliding hockey puck being kicked or an orbiting rocket turning on its thrusters (Clement, 1982; Kozhevnikov et al., 2007). Although these types of problems are not technically projectile motion problems because another force besides gravity is acting on the object, they are grouped with projectile motion problems because they require participants to combine two vectors of motion to predict an object's path.

Projectile motion problems are challenging for many participants, whose answers tend to fall into one of several categories. For example, when asked to predict the path of a ball after it is released from a flying plane, 60% of participants provide incorrect responses including a straight diagonal line, a backwards trajectory, and a path straight down (McCloskey, 1983b). (The correct answer is a forward, parabolic curve because the object maintains its horizontal velocity after being released, while simultaneously accelerating downward from gravity.) Errors are both frequent and consistent: Participants provide similar patterns of incorrect responses for other projectile motion problems (Caramazza et al., 1981; Champagne, Klopfer, & Anderson, 1980; Kozhevnikov et al., 2007; McCloskey et al., 1983).

Similarly, curvilinear motion problems (movement along a circular path) are challenging for participants. Curvilinear motion problems involve an object traveling in a curved trajectory due to a moving vehicle then being released at a specified point along its trajectory such as a ball rolling out of a curved tube (McCloskey et al., 1980), a ball breaking from a string as it is twirled above a person's head (McCloskey & Kohl, 1983) or water running out of a coiled garden hose (Kaiser et al., 1986). When presented with a curved tube problem, 33% of participants predict the ball would continue in a curved path once it exited a curved tube (McCloskey et al., 1980). (The correct answer is that the ball would continue moving in a straight path because once it exists the confines of the tube, the tube is no longer exerting a force on the ball and guiding its movement.) Again, errors are consistent: This incorrect prediction appears in 50% of responses for participants presented with a spiral tube problem, 30% of responses for participants presented with the twirling ball and string problem (McCloskey et al., 1980), and 37% of responses for participants presented with the garden hose problem (Kaiser et al., 1986).

Past experience does not necessarily bolster kinematic problem solving, attesting to the difficulty of kinematic reasoning for many adults. For example, participants with previous physics education at the high school, college, or graduate level still produce incorrect predictions despite having explicit knowledge about how objects interact with forces (Caramazza et al., 1981; Kozhevnikov & Hegarty, 2001; McCloskey et al., 1980). Although formal physics education can lead to fewer errors, adults with high school or college level education in physics make the same type of errors as those with less education and produce 48% of incorrect responses (McCloskey, 1983b; McCloskey et al., 1980). Additionally, experience with a subset of kinematic problems, such as projectile motion, does not necessarily improve competency with other kinematic problems, such as curvilinear. Experience with daily kinematic problems does not transfer to abstract, but equal, kinematic problems like those presented in physics courses. For example, 66% of participants correctly predict the flow of water out of a hose (familiar problem) while only 39% of participants correctly predict the path of a ball exiting the same hose (abstract problem) (Kaiser et al., 1986). When results from the participants who solved the water hose problem first were compared to results from the participants who solved the ball problem first, there was no significant difference. These findings suggest that it is difficult to generalize existing kinematic knowledge to abstract, novel problems (Kaiser et al., 1986) perhaps because adults are not identifying similar characteristics between different problems (Holyoak, Gentner, & Kokinov, 2001).

Theories About Kinematic Knowledge Acquisition

How do we acquire kinematic knowledge and what does it reveal about the frequency and the characteristics of kinematic errors? Three hypotheses—seeing-is-believing, impetus, and action-on-objects—attempt to answer this question. The hypotheses share a focus on how visual and haptic information lead to misinformation, but they differ on the explanations for object motion abstracted from the misinformation.

The seeing-is-believing hypothesis posits that misconceptions about object motion develop because the perception of an object motion event does not reflect unbiased reality. In other words, we experience perceptual illusions that distort how we think objects move (McCloskey, 1983a, 1983b; McCloskey & Kaiser, 1984; McCloskey et al., 1983). Perceptual illusions are thought to be influenced by the viewer's frame of reference during an object motion event (McCloskey & Kaiser, 1984). Consider a plane dropping a package and a viewer on the ground observing the event. If the viewer looks up and maintains the plane and package in their sight, the plane becomes the frame of reference for how the package is falling: The package remains in line with the plane and appears to fall straight down. This illusion disappears if the frame of reference changes from the plane to the ground: If the viewer watches the event with the ground in his field of view, he can observe the package at the point of release and see its entire parabolic trajectory. Additionally, if the viewer switches between the two frames of reference, alternative perspectives emerge. For example, people predict the package moves horizontally for some time after it is released then falls straight down, moving in an inverse L-shaped path. Another prediction indicates the package will fall forward and downward at the same rate, moving in a diagonally forward path. Switching between frames of reference enables a person to observe the package falling straight down relative to the plane and forward relative to the ground, creating an opportunity to combine the vectors of motion is different configurations to arrive at an answer. The seeing-is-believing hypothesis provides an explanation for some commonly made projectile motion errors, such as the straight down, diagonal, and inverse L-shaped responses (McCloskey et al., 1983).

However, the seeing-is-believing hypothesis cannot explain the backwards parabolic predictions for projectile motion and the incorrect predictions for curvilinear motion. Instead, such errors are better explained by the impetus theory, which proposes that observers develop a sophisticated set of motion beliefs based on witnessing kinematic events under various circumstances. Kozhevnikov and Hegarty (2001) summarize the features of the impetus theory in five beliefs: (1) objects acquire impetus (an internal force) when acted on by other objects, (2) objects that are simply dropped do not acquire an impetus because another object is not acting on it, (3) an impetus will gradually dissipate, (4) objects traveling through air will only be subjected to gravity once its impetus has dissipated and (5) there are different types of impetus (linear and curvilinear).

The impetus theory is based on interviews and think-aloud procedures using kinematic problems. For example, when asked to describe why a ball would follow the predicted path of motion, 85% of participants report that the ball would acquire a “force”, “momentum”, or “something” that would “keep it in motion”. Participants who indicate that a ball would continue in an inverse L-shaped path after rolling off a cliff explain that the ball would slow down after a period of time after rolling off the cliff ledge because it would “lose force” and “gravity would take over” (McCloskey, 1983b). Similar explanations were obtained in other studies (Cook & Breedin, 1994; McCloskey & Kohl, 1983), revealing a systematic—even if incorrect—set of rules to explain how an impetus accounts for an object’s motion under various conditions.

The first two theories, seeing-is-believing and impetus, focus primarily on visual experiences. However, kinematic experiences are not exclusively visual; they are also composed of haptic interactions such as throwing a ball or driving a car. The actions-on-objects hypothesis states that our understanding of forces and causal events, such as object motion, are the result of daily physical interactions with objects and the development of a general heuristic about what properties moving objects can impart onto other objects (White, 2009, 2012, 2013). Before acting on an object, people create an internal representation of where an object will go and how much force is required to get it there. Once an object is physically acted on, like pushing a bowling ball, people gather force information from sensory receptors in their skin. After the action is executed, people compare their internal representations with the outcome and sensory information. The comparison process creates a feedback loop that updates existing internal representations of object motion and force. Imagine pushing a bowling ball and a golf ball across the ground. Moving the bowling ball requires more effort than moving the golf ball because the bowling ball is heavier. Even when the bowling ball starts moving, it does not travel as far as the golf ball when equal effort is applied. From this experience, we gather that heavier objects require more effort to move and the force appears to lessen as the object moves further away. Additionally, we learn that certain properties of a force, like directionality, can be transferred to an object.

Unfortunately, generalizing from previous experiences can lead to kinematic errors. For example, when told to imagine that a hockey puck is sliding from left to right on ice and we kick it from a perpendicular direction, 34% of adults believe it will follow the direction of the force (perpendicular) while only 44% of adults correctly predict it will move along a forward diagonal path (Kozhevnikov et al., 2007). When presented with a pendulum problem and asked to predict the path of the ball if detached from the string while swinging, 75% of participants incorrectly predict the ball will fall straight-down, parabolic backwards, or in an inverse L-shaped path (Caramazza et al., 1981), perhaps because they are referencing multiple previous experiences to solve the problem and ascribing the force to different parts of the scenario. If force is ascribed to the ball, the weight of the ball would be the primary influencer of its direction. A heavy ball would be predicted to fall straight down, or resist swinging forward and fall backwards. A lighter ball might follow its forward direction once it is released but then fall straight down due to gravity or fall along a forward path. Other adults who assign force to the swinging string may use its position to guide their judgement. Therefore, depending on where the string is positioned, their prediction changes (Caramazza et al., 1981). Misunderstanding force and object motion under different circumstances also explains the curved path responses for the curvilinear problems. Assigning the mechanism of force to the curved tube would lead adults to predict that the ball would follow the directionality of the tube and depending on the perceived weight of the ball, the angle of curvature would vary. The actions-on-objects hypothesis accounts for the pattern of responses seen in projectile motion problems, explaining the backward response not accounted for in the previous hypotheses, as well as the pattern of curvilinear problem responses.

A Movement for Improvement

All three hypotheses—seeing-is-believing, impetus, and action-on-objects— share an important characteristic: Adults create incorrect mental representations of spatial relations and use the representations to guide their predictions. The key to improving kinematic problem solving may lie in improving spatial visualization, a spatial ability strongly correlated with performance on projectile motion problems (Kozhevnikov et al., 2002). This relationship is not surprising: Kinematic problems in real life and in the laboratory often involve picturing—or the spatial visualization of—object movement. Indeed, high spatial visualizers (participants who score in the top 25% of a spatial ability score distribution) outperform low spatial visualizers (those in the bottom 25%) on several multiple-choice kinematic problems (Kozhevnikov et al., 2007). The high spatial visualizers often accurately identify both the horizontal and vertical components of motion and integrate them correctly into their problem-solving strategies. In contrast, low spatial visualizers tend to focus on a single component of motion such as horizontal or vertical.

Spatially-oriented language provides helpful hints about how people think and communicate about kinematic problems. When participants were asked to explain their reasoning for their responses to a kinematic problem set, 43% of errors were due to omitting an important component of the problem, such as the horizontal velocity, gravity, thrust, or acceleration (Cook & Breedin, 1994). Interview responses indicate that low spatial visualizers are more likely to focus on object characteristics such as weight, speed, and internal force, whereas high spatial visualizers are more likely to focus on how external forces would influence an object's path

(Kozhevnikov et al., 2007). In other words, high spatial visualizers are more likely to focus on the spatial relations of an object (e.g., an object’s direction) compared to low spatial visualizers, who focus more on the object (e.g., an object’s weight).

Additionally, improvements in spatial visualization have led to improvements in other spatial abilities and kinematic problem solving. Spatial training studies that provide participants with spatial visualization tasks, such as the paper folding task and the form board task, demonstrate improved performance on other spatial visualization tasks as well as mental rotation tasks, such as the Cube Comparison Test and the Shepard and Metzler Rotation Task (Uttal et al., 2013). Additionally, improvements in spatial visualization coincide with improvements in kinematic problem solving when using a computer-based learning program and assessing pretest-to-posttest gain scores (Kozhevnikov & Garcia, 2011). Increased performance on spatial and kinematic tasks suggests that improvements in spatial visualization can improve performance on similar kinematic problems (near transfer tasks) and spatial problems that use the same spatial abilities (far transfer tasks).

Spatial visualization utilizes visual-spatial working memory (Isaak & Just, 1995). Baddeley and Lieberman’s (1980) working memory model includes a visual-spatial subsystem equipped with a visual-spatial sketchpad where we can imagine and manipulate mental images, possibly through the use of a cognitive coordinate system (Just & Carpenter, 1985; McCloskey, Valtonen, & Cohen Sherman, 2006; Salthouse, Babcock, Mitchell, Palmon, & Skovronek, 1990). A coordinate system is one way to track an object’s position or internal object changes such as folding, and can change depending on the demands of the task (Just & Carpenter, 1985). With respect to kinematic problems, the most helpful way to represent object and spatial locations is to mentally superimpose a grid on the entire problem to track the transformations over time, as evidenced by the time-related spatial language provided by high spatial visualizers (Kozhevnikov et al., 2007).

Picturing spatial relations and tracking changes is challenging and people struggle with one or more steps of the process because the visual-spatial working memory system can become overloaded if required to hold and process too much information (Salthouse, Babcock, Mitchell, et al., 1990; Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990; Salthouse, Mitchell, Skovronek, & Babcock, 1989). Therefore, deficits in kinematic reasoning could be the result of people defaulting to misconceptions about object motion created from perceptual misinformation, as discussed in each of the three hypotheses, because it is too difficult to picture the object motion event. Decreasing the visuo-spatial cognitive load would give adults an opportunity to mentally work through a problem, and we should observe improvements in adult's kinematic reasoning.

Peer Collaboration

Working with a peer, or peer collaboration, can create an opportunity to decrease individual cognitive load. A partner can identify components of a problem not being considered or suggest an alternative perspective which can be incorporated into an existing perspective (Vygotsky, 1978). For example, when predicting where a ball will fall after rolling off a table, one person might predict it falling straight down next to the table leg. The other person might point out that the ball is rolling forward, not being dropped, and should continue moving forward. After discussing the possibilities, the peer collaborators might correctly conclude that the ball will land further away from the table because the horizontal movement of the ball is conserved.

The exchange of ideas between peers can lead to improved conceptual knowledge and reasoning skills, as evidenced by pretest-posttest gain scores for both children and adults who worked in dyads (Azmitia, 1988; Dimant & Bearison, 1991; Fawcett & Garton, 2005; Kozhevnikov & Thornton, 2006; Phelps & Damon, 1989). Conversation and the exchange of new perspectives is believed to facilitate the improvements after evaluating the performance of participants who either worked alone, worked alone and talked out loud, or worked with a peer (Fawcett & Garton, 2005; Teasley, 1995). More importantly, improvements in reasoning skills after working with a peer persists for spatial learning. For fourth-graders who worked in pairs for one year on either mathematics or spatial reasoning concepts, dyads in the spatial reasoning groups demonstrated significantly higher improvements between pre-test and post-test scores than the mathematics and control groups; improvements were still apparent one year later (Phelps & Damon, 1989). Five-year-old children tasked with reconstructing a LEGO building were more successful at copying the model when they worked with a peer than those who worked independently (Azmitia, 1988). Additionally, over the course of a semester, college students demonstrated increased spatial visualization ability and physics competency after given the opportunity to meet with a peer at the end of each class to discuss the lesson and solve computer-based physics problems (Kozhevnikov & Thornton, 2006). Overall, working with a peer and discussing ideas and problem-solving strategies appears improve spatial ability and deepen conceptual knowledge.

However, there are limitations if certain criteria are not met. The composition of the dyads and quality of discussion appear to play a large role. There was no significant difference in performance on a Piagetian spatial task between mixed-gender dyads, same-gender dyads, and an individual condition, and is thought to be the result of unsubstantial conservation about the task demands and alternative solutions (Golbeck & Sinagra, 2000). Drawing comparisons between this study and previous literature reveals that the quality of interaction, the length of time working together, and the type of task influence the positive effects of peer collaboration. Golbeck and Sinagra noted that the dialogue between peers lacked focus on the task and discussion about their individual approaches to the problem. Alternatively, the results could reflect the task being inappropriate for collaboration because it relies on an individual's kinesthetic sense of horizontality rather than a mental representation (Golbeck & Sinagra, 2000) or requires less conceptual reasoning necessary to promote discussion (Phelps & Damon, 1989). Another explanation is that both partners did not have different perspectives on the problem and reinforced existing perspectives and reasoning. Additionally, previous studies took place over the course of several weeks or months while the study conducted by Golbeck and Sinagra (2000) took place in one session. Results suggest that the effect of peer collaboration takes longer to produce conceptual change and improved spatial ability. Additionally, it is possible that because the discussion lacked focus on task demands, limited use of spatial language and task-related discussion failed to prompt greater spatial reasoning in the participants. There was a significant difference in performance on a computer-based spatial task between fourth graders working alone, working alone but talking out loud, and working in dyads. However, the amount of discussion about the problem was a greater predictor of success on the task than the experimental condition (Teasley, 1995). Students in the alone talk and collaborative conditions generated more hypotheses and used more descriptive language than students in the alone condition. However, students in dyads spent more time discussing the computer-program and assessing outcomes, demonstrating a deeper level of evaluation than students in the alone conditions (Teasley, 1995). This suggests that the quality of discussion and use of spatial language or language directly related to the task demands is an important element of working with a peer. Therefore, the quality of discussion, specifically, the type of words being used amongst partners, should be considered when assessing the effect of peer collaboration on kinematic problem solving to provide insight into the ideas being exchanged between partners.

Current Study

The goal of this study was to investigate whether peer collaboration is an effective tool for improving kinematic problem-solving skills. Participants in the key experimental condition practiced solving kinematic problems collaboratively while talking through the problems together. Additional participants solved the same problems alone while talking out loud (a condition that served as a control for the effects of collaborative learning) or alone without talking out loud (a condition that served as a control for both the effects of collaborative learning and time spent talking out loud). Afterwards, all participants completed transfer tasks to determine whether collaborative learning improved kinematic problem solving at all (near transfer task) and whether it had far-reaching consequences (far transfer task).

The practice kinematic problems included the airplane problem, cliff problem, pendulum problem, ramp problem, and the conveyor belt problem (Caramazza et al., 1981; Kozhevnikov et al., 2007; McCloskey, 1983b). These problems are widely used in projectile motion research. Therefore, results were evaluated using criteria from previous studies. The near transfer problem set consisted of the walker problem, the hockey puck problem, and the rocket problem. These problems were comparable to the kinematic practice problem set because they involved utilizing the same components of motion (horizontal and vertical) but required the participants to organize the information in a different spatial context and integrate the components in a new configuration. The far transfer problem set consisted of paper folding and cube comparison problems. These problems are widely used to assess spatial visualization ability and mental rotation ability (Ekstrom et al., 1976; Linn & Petersen, 1985; Uttal et al., 2013), abilities necessary for solving kinematic problems. Additionally, particular attention was paid to the amount and type of language participants used to communicate their ideas. Similarly, we examined the potential effects of other variables shown to influence kinematic skills and spatial problem solving (gender, duration of practice, physics education, spatial experience, and individual spatial ability).

First, we predicted that participants who practiced with a peer would perform better on the kinematic practice problems compared to participants who practiced alone, possibly due to discussing task-related demands and exchanging alternative solutions. Second, we predicted that participants who practiced kinematic problems with a peer would perform better on near and far transfer tasks compared to participants who practiced alone. We expected that the quality of discussion, specifically discussion of spatial relations, during the practice portion would facilitate more accurate spatial visualization and enable participants in the Collaborative condition to perform better on transfer tasks compared to participants in the Alone conditions. Lastly, we predicted that participants who practice kinematic problems with a peer would use a greater amount of language, specifically, a greater amount of spatial language as a result of discussing task-related demands and solutions with a partner compared to participants who practiced alone.

Method

Participants

A power analysis using G*Power showed that for a repeated-measures ANOVA with three groups, two measurements, and a 0.05 level of significance, a sample size of 66 participants would be necessary for a Cohen’s *d* = .02 and a power level = 0.80. Since the key experimental condition (Collaborative condition) requires two people per session, the final sample size needed to be equal to or greater than 88 participants: 22 individual participants per Alone condition and 22 dyads (44 individual participants) in the Collaborative condition.

Ninety-nine college-age adults participated for partial credit toward a course research requirement. Participants were 28 men and 69 women, with one participant opting to not indicate their gender. Participants were between 18.27 and 23.14 years of age (*M* = 19.32). Nine participants (3 men, 6 women) were excluded from the final sample due to technical difficulties with the recording set up (*n* = 2) and failing to follow instructions (*n* = 7), leaving a final sample size of ninety participants.

Design

We used a repeated measures design with five covariates. The independent variables were practice condition and transfer task. Practice condition was defined by the number of people working on the practice problem set and whether they were instructed to talk during the practice portion. There were three levels: Alone-Quiet (one person quietly completing the practice problems), Alone – Talk (one person completing the practice problems while talking out loud through his or her problem solving), and Collaborative (two people working together to complete the practice problems). Transfer task was defined as a problem set that required participants to use similar spatial reasoning and abilities. There were three levels: kinematic near transfer task (kinematic problems similar to the practice problems), paper-folding far transfer task (spatial task that requires spatial visualization ability), and cube comparison far transfer task (spatial task that requires mental rotation ability). Twenty two participants were assigned to each of the two control conditions (Alone-Quiet and Alone-Talk practice conditions) and 46 participants (23 dyads) were assigned to the experimental condition (Collaborative practice condition). Participants were assigned randomly to one of the three practice conditions as they signed up for a session. Effort was made to counterbalance the number of women and men in each condition. All participants individually completed the transfer tasks.

The dependent variable was accuracy on transfer task, defined as the proportion of correct responses to completed responses. The covariates were physics education (the number of physics courses completed), spatial experience (the frequency each individual participated in spatial activities), individual spatial ability (the proportion of correct responses to completed problems on the Spatial Ability Test), time (duration of time an individual or dyad spent working on the kinematic practice problem set), and gender (as indicated on the demographic questionnaire). Additionally, spatial language was explored separately for quantity (frequency of utterances) and quality (type of utterances).

Materials

All sessions were videotaped using a Canon Vixia HF R52 HD Camcorder. Datavyu open-source software (Version v1.4.1; Datavyu Team, 2014) was used to code video and audio data.

A demographic questionnaire collected general participant information including birthdate, test date, and gender. Physics education was measured by asking participants to indicate their highest level of physics education and the number of physics courses they completed. The physics education questionnaire was adapted from McCloskey and Kohl (1983). Spatial experience was measured using an 81 item Spatial Activity Questionnaire (Baenninger & Newcombe, 1989). Each item consisted of an activity (e.g., soccer) followed by a 6-point Likert scale (never participated, participated less than 4 times, participated 5-15 times, participated in about once a month, participated in about once a week, participated in more than once a week) to indicate the frequency of participation between 13 and 18 years of age. Spatial ability was assessed with a Spatial Ability Test consisting of 20 questions pulled from the Manual for Kit of Factor-Referenced Cognitive Tests that include the Card Rotations Test, Form Board Test, and Surface Development Test (Ekstrom et al., 1976). The Card Rotations Test present one figure followed by five similar figures. The similar figures could either be identical figures but rotated clockwise or counterclockwise or mirror images rotated clockwise or counterclockwise. Participants were instructed to only circle the identical images. The Form Board Test presented an original shape and two groups of different shapes. Participants were instructed to select the group of shapes that would create the original shape if assembled together. The Surface Development Test presented a folded cube with three visible faces each with a different letter, and four deconstructed cubes each with the same letters in different positions. Participants were instructed to select the deconstructed cube that matched the folded cube. The kinematic practice problem set consisted of five widely-used, paper-and-pencil projectile motion problems: the airplane problem, cliff problem, pendulum problem, conveyor belt problem, and ramp problem (Caramazza et al., 1981; McCloskey, 1983b). Each problem depicted a different scenario of a ball in motion. Participants were asked to indicate the path the ball would travel in each scenario by drawing the path with a pencil. The near transfer assessment consisted of three paper-and-pencil kinematic problems: the hockey puck problem, walker problem, and rocket problem

(Caramazza et al., 1981; Kozhevnikov et al., 2007). Each problem depicted a different scenario of an object in motion (e.g., ball, hockey puck, rocket). Participants were asked to indicate the path each object would follow in each scenario. Far transfer was assessed using two spatial tasks: paper-folding and cube comparison (Ekstrom et al., 1976). The Paper-Folding task consisted of ten items with each item showing a successive drawing of three folds made to a square sheet of paper. The final drawing in the series was a folded paper with holes punched through it. Participants were instructed to select one drawing from five drawings that showed how the punched sheet of paper would look when fully opened. The Cube Comparison task consisted of 21 items with each item consisting of a drawing of two cubes with letters printed on each side. Participants judged whether the two cubes were of the same cube or different cubes.

Procedure

At the start of the session, an experimenter explained the purpose of the study and guided participants through the informed consent process. Next, the experimenter turned on the video camera and began recording the session. All participants completed the following questionnaires individually, one at a time, and in the following order: demographics (untimed), physics education (untimed), Spatial Activity Questionnaire (untimed), and Spatial Ability Test (10-minute limit).

The practice phase (untimed) began immediately following the questionnaires. The experimenter provided participants with a booklet of practice problems. The participants in the Alone-Quiet condition received the following verbal instructions from the experimenter: “Please complete the following five problems in order. Indicate your answer by drawing your answer on the paper. Please mark your answers clearly with a pencil and completely erase any extra lines. You may take as long as you need to complete these problems. When you are finished with all five problems, please notify me.” The participants in the Alone-Talk condition received the following verbal instructions from the experimenter: “Please complete the following five problems in order. Indicate your answer by drawing your answer on the paper. As you complete the problem, explain your reasoning out loud and how you arrived at your answer as if explaining the problem to a classmate. That is, please talk through your thinking and reasoning out loud. Say what you’re thinking. Please mark your answers clearly with a pencil and completely erase any extra lines. You may take as long as you need to complete these problems. When you are finished, please notify me.” The dyads in the Collaborative condition received the following instructions from the experimenter: “Please complete the following five problems in order by working together. Indicate your answer by drawing your answer on the paper. As you work through the problems, please explain your reasoning and how you arrived at your answer to your partner. That is, talk to each other about your reasoning. You can only provide one answer per problem. If you do not agree on an answer, please discuss the problem until you come to an agreement. Please mark your answers clearly with a pencil and completely erase any extra lines. You may take as long as you need to complete these problems. When you are finished, please notify me.”

After the practice phase, each participant completed the assessment phase independently. Participants received the near transfer kinematic assessment first (untimed). Then, participants received the Paper-Folding far transfer assessment (3-minute limit). Next, participants received the Cube Comparison far transfer assessment (3-minute limit). Finally, the experimenter debriefed each participant to conclude the session and turned off the video camera.

Scoring and Coding

Near and Far Transfer Assessments. Responses to the near transfer kinematic problem set were assessed for accuracy using the criteria from previous studies (i.e., Caramazza, McCloskey & Green, 1981; McCloskey, 1983; McCloskey, Washburn & Felch, 1983; Kozhevnikov, Motes & Hegarty, 2007). Each participant received a score based on the number of correct responses and this score was converted into a proportion using the number of correct responses divided by the total number of problems. Responses to the far transfer assessment were assessed for correctness and each participant received an individual score for the paper-folding task and the cube comparison task. Considering the total number of problems for each task is different, we standardized the scores by using a proportion, the number of correct responses divided by the number of completed problems, for each far transfer assessment. To make sure participants' effort was valid, we calculated the average number of correct responses for the paper-folding task (*M* = 4.54, *SD* = 1.94) and the cube comparison task (*M* = 14.92, *SD* = 4.15) and the average number of completed problems for the paper-folding task (*M* = 6.86, *SD* = 1.85) and the cube comparison task (*M* = 17.16, *SD* = 4.06). The distributions were similar to previous studies that used these tasks (Kozhevnikov et al., 2002; Kozhevnikov et al., 2007), therefore, using a proportion was determined to be a valid measurement of accuracy.

Language. Using Datavyu software, each data session was reviewed by both primary and secondary coders to mark the beginning and end of each task (e.g., questionnaires, practice problem set, transfer assessments). Individual trials were delineated for the kinematic practice problem set, as this was the task of interest for spatially-oriented language use. For each task or trial, a primary coder marked the start of each utterance using the typical rule based on pause and content (Dancu, Gutwill, & Sindorf, 2009; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Pruden, Levine, & Huttenlocher, 2011; Teasley, 1995). An utterance was defined as a word or phrase separated by a 2-3 second pause from another verbalization, or a word or phrase communicating the same topic (Cannon, Levine, & Huttenlocher, 2007). Each utterance was coded as one of 11 language categories, which are described in detail in Table 1. The categories were modeled after previous language coding manuals on spatial language and coding schemes for spatial problem solving (Cannon, Levine, & Huttenlocher, 2007; Teasley, 1995; Winsler, Fernyhough, McClaren, & Way, 2005).

Table 1

Language Coding Categories

Utterance Description	Examples	Reliability (*r*)
References to other problems in the problem set: indicates that the current problem is similar or different to a previous problem in the experiment	"This is like the airplane problem."	.92
Personal experience: refers to a personal experience that is related to the problem	"I remember a problem like this from high school."	NA
External example: provides an example of an event that is similar and/or different from the current problem that serves as a way to understand the current problem better	"It's like going down a slide."	.68
Visualization prompt: words or phrases that help the participant mentally represent the problem, mentally viewing the problem from a specific perspective, or picture themselves enacting the task	"Imagine", "If you look at it from the side", "If you think about it...", "If I were dropping the ball, I think I would drop it earlier..."	NA
Scientific/physics related concepts: words or phrases used to convey scientific or physics concepts	momentum, gravity, force, speed, velocity	.77
Direction of motion: where the object is moving	forward, backward, straight-down, up, down, left, right	.76
Manner of motion: how the object is moving	spinning, rolling, falling, gliding, curving, arc	.50
Location in space: where the object is in space or relative to another object in the problem	above, on top, below	-.09
Object description: words that describe how an object looks, feels, or functions	big, round, metal, steep, "The conveyor belt holds this…"	.40
General movement: conveys motion but not a specific direction or use of a specific verb	this way, going, "go like that", "it will hit the ground"	.71
Unintelligible/other: utterances that are inaudible, uninterpretable, or did not fit any other category	"Yea", "Okay"	.86
Total number of utterances		.90

A second coder coded 20% of the kinematic practice problem trials and 33% of the transfer assessment trials for inter-rater reliability. Because the language coding and analyses were largely exploratory, we used a conservative method to calculate inter-rater reliability, using only the trials that received at least one language code, either from the primary or secondary coder. Trials where the primary and secondary coder agreed no language occurred were excluded from the reliability calculation, even though these instances were technically in agreement. The two coders had high reliability on the type of task/trials and whether language occurred at all on those tasks/trials (r = .99) and total number of utterances (r = .90). Reliability ranged from low to high on the number of utterances for each category (r = -.09 - .92[1]). It was not possible to calculate reliability for two of the language categories, visualization prompts and personal experiences, due to the low number of occurrences in these categories.

---

[1] Some of the lower reliability rates were due to the low number of trials in which the participants used this category. For example, the location category only contained 12 data points (r = -.09).

Results

The primary goal of this study was to examine whether practicing kinematic problems with a peer improved kinematic problem solving and, by extension, related spatial abilities. Additionally, we were interested in how language facilitated spatial reasoning. The practice problems and transfer tasks were analyzed separately to explore what happened during initial learning and whether it may have influenced transfer.

Kinematic Practice Problems

Of the five practice problems, participants across the three conditions averaged 1.49 correct responses (*SD* = 1.04, mean proportion = .30), attesting to their difficulty with kinematic problems. This finding is consistent with results from past studies (e.g., Caramazza, McCloskey & Green, 1981; McCloskey, 1983; McCloskey, Washburn & Felch, 1983; Kozhevnikov, Motes & Hegarty, 2007). Refer to Figure 1 for the following results. Contrary to our prediction, participants who practiced with a peer did not perform better on the kinematic practice problems compared to participants who practiced alone. A one-way ANCOVA on the practice conditions (Alone-Quiet, Alone-Talk, Collaborative) with five covariates (gender, duration of practice, physics education, spatial experience, and spatial ability, described further in Table 2) showed no effect of practice condition, $F(2, 81) = 0.90$, $p = .41$, $\eta_p^2 = .02$.

Table 2

Descriptive Statistics for Continuous Covariates (N = 90)

Covariate	Minimum	Maximum	Mean	SD
Duration of practice (seconds)	72.37	279.20	134.20	41.94
Physics education (number of classes)	0	5	0.88	0.68
Spatial experience score (proportion)	0.22	0.42	0.29	0.05
Spatial ability score (proportion)	0.05	0.95	0.55	0.22

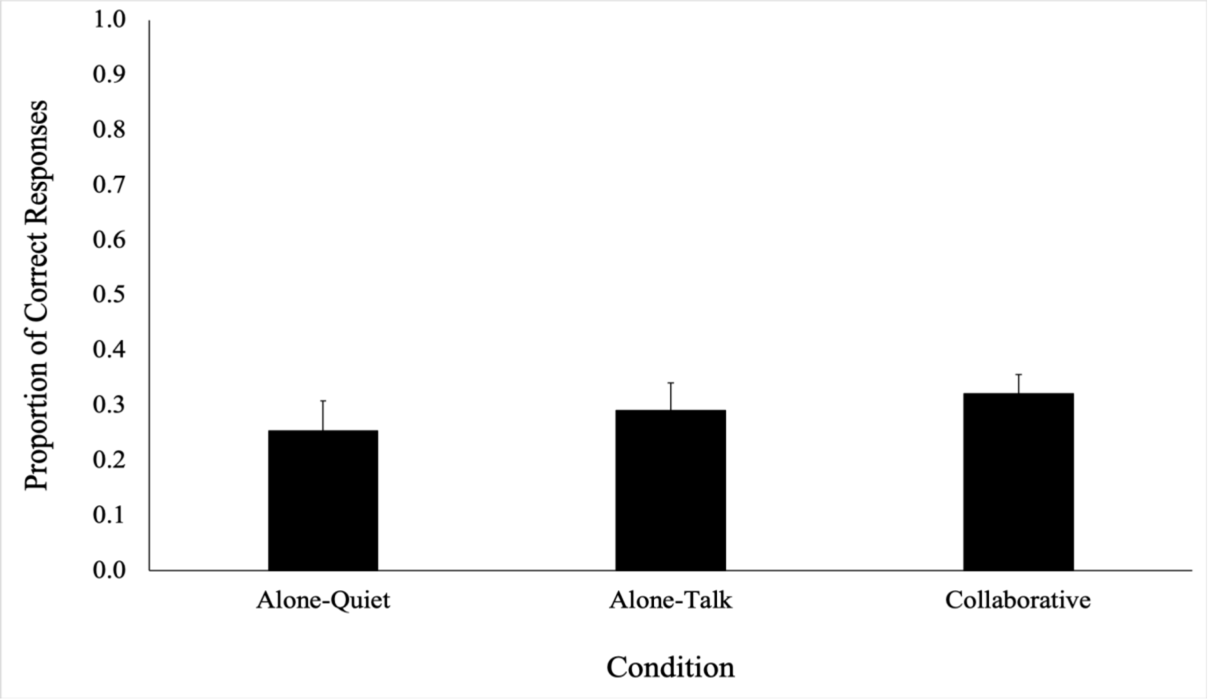

Figure 1. Accuracy on the kinematic practice problem set for each practice condition. Error bars indicate mean standard error.

Dyads in the Collaborative condition (*M* = 0.32, *SD* = 0.17) did not perform better—or worse—compared to individuals in the Alone-Quiet (*M* = 0.25, *SD* = 0.26, *p* = .77, *d* = 0.32) and Alone-Talk (*M* = 0.29, *SD* = 0.24, *p* = .84, *d* = 0.14) conditions. Duration of practice, physics education, spatial experience, and individual spatial ability were not significant predictors of accuracy on the kinematic practice problem set (*p*s > .18). However, gender was significant (*p* = .01, η_p^2 = .08). (See Table 3 for descriptive statistics for gender.) Follow up *t*-tests using Bonferroni correction were conducted within each condition with gender as the independent variable and proportion of correct responses on the kinematic practice problem set as the dependent variable. With the Bonferroni correction, there was no significant difference in performance between men and women in the Alone-Quiet, *t*(20) = 2.18, *p* = .04, *d* = 0.87, Alone-Talk, *t*(19) = 2.81, *p* = .35, *d* = .39, or Collaborative condition, *t*(44) = 1.84, *p* = .07, *d* = .39.

Taken together, these results suggest that practicing with a peer did not influence kinematic problem solving.

Table 3

Practice Problem Scores for Men and Women

	Gender									
	Men					Women				
Practice Condition	*n*	Mean	SD	Min.	Max.	*n*	Mean	SD	Min.	Max.
Alone – Quiet	6	.43	.34	0	1	16	.19	.19	0	.60
Alone – Talk	8	.35	.30	0	.80	13	.25	.20	0	.60
Collaborative	11	.40	.18	0	.80	35	.30	.16	0	.80

Table 4

Descriptive Statistics for Language

	Total Language					Content Language				
Practice Condition	*n*	Mean	*SD*	Min.	Max.	*n*	Mean	*SD*	Min.	Max.
Alone – Quiet	22	2.82	3.69	0	13	22	0.09	0.43	0	2
Alone – Talk	22	21.18	10.84	8	46	22	11.18	5.85	3	26
Collaborative (dyads)	23	64.74	48.05	24	237	23	26.83	18.26	5	83

Although practice condition did not affect performance, we continued to explore spatial language use in case it influenced transfer task performance. Table 4 presents the descriptive information for language within each practice condition. A one-way ANCOVA was conducted using the total number of utterances as the dependent variable and practice condition as the independent variable, whiling controlling for duration of practice (see Figure 2a). As expected, condition had a significant effect on the amount of language used, $F(2, 63) = 46.39$, $p < .001$, $\eta_p^2 = .60$, and duration of practice was a significant covariate ($p < .001$, $n_p^2 = .26$).

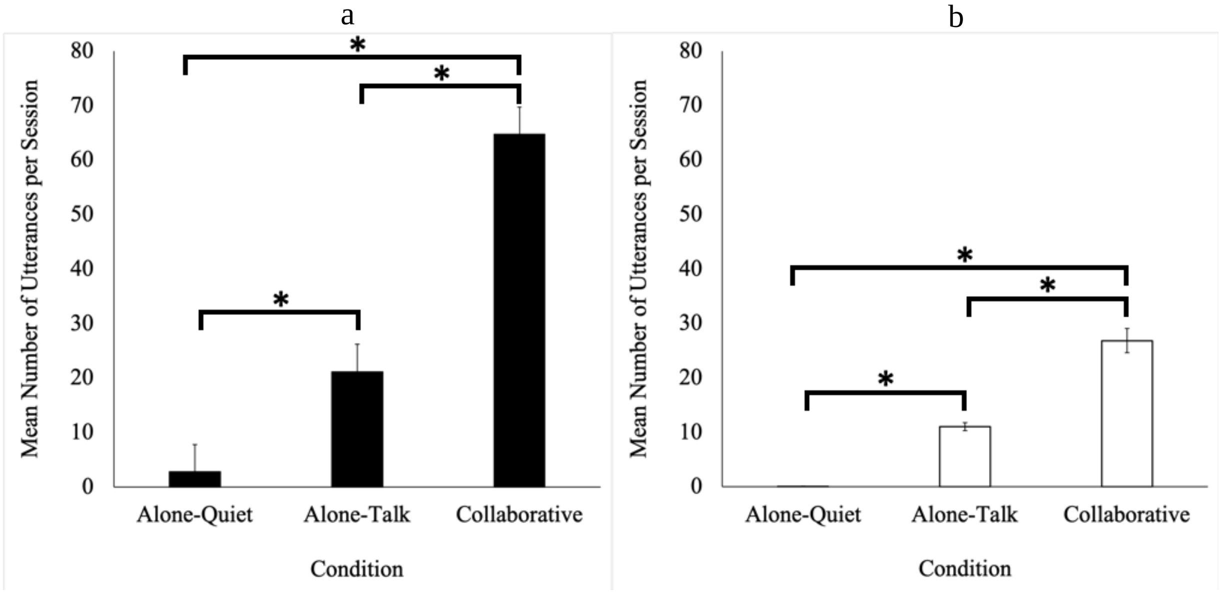

Figure 1. (a) Mean number of utterances (including the "unintelligible/other" category) per session for each practice condition. (b) Mean number of utterances (excluding the "unintelligible/other" category) per session for each practice condition. Error bars indicate mean standard error.

However, Levene's test for equality of variances was significant (p = .002), indicating that the variance within each practice condition was not equal. Therefore, a Kruskal-Wallis non-parametric test was used to confirm the results of the ANCOVA. The Kruskal-Wallis H test confirmed condition-related differences in the amount of utterances, $\chi^2(2) = 53.62$, $p < .001$. Dunn's post hoc tests with a Bonferroni correction indicated the Collaborative condition used significantly more language than the Alone – Talk ($p < .001$, $d = 1.25$) and the Alone – Quiet conditions ($p < .001$, $d = 1.82$), and the Alone – Talk condition used significantly more language than the Alone – Quiet condition ($p < .001$, $d = 2.27$). As depicted in Figure 2b, when irrelevant content (utterances coded as "unintelligible/other") was removed from analysis, practice condition continued to exert a significant effect on language usage, $F(2, 63) = 44.53$, $p < .001$, $\eta_p^2 = .59$, and duration of practice remained a significant covariate ($p < .001$, $\eta_p^2 = .17$).

However, Levene's test for equality of variances was significant ($p < .001$). A Kruskal-Wallis non-parametric test was used to confirm the results of the ANCOVA. The Kruskal-Wallis H test confirmed condition-related differences in the amount of content utterances, $\chi^2(2) = 52.04$, $p < .001$. Dunn's post hoc tests with a Bonferroni correction indicated the Collaborative condition used significantly more content language than the Alone – Talk ($p = .013$, $d = 1.15$) and the Alone – Quiet conditions ($p < .001$, $d = 2.07$), and the Alone – Talk condition used significantly more language than the Alone – Quiet condition ($p < .001$, $d = 2.67$). Thus, even though participants showed similar accuracy on the practice problems, there were condition-related differences in how much participants talked and how much they used relevant language. However, this difference, in part, can be attributed to the Collaborative condition spending more time on the practice problem set.

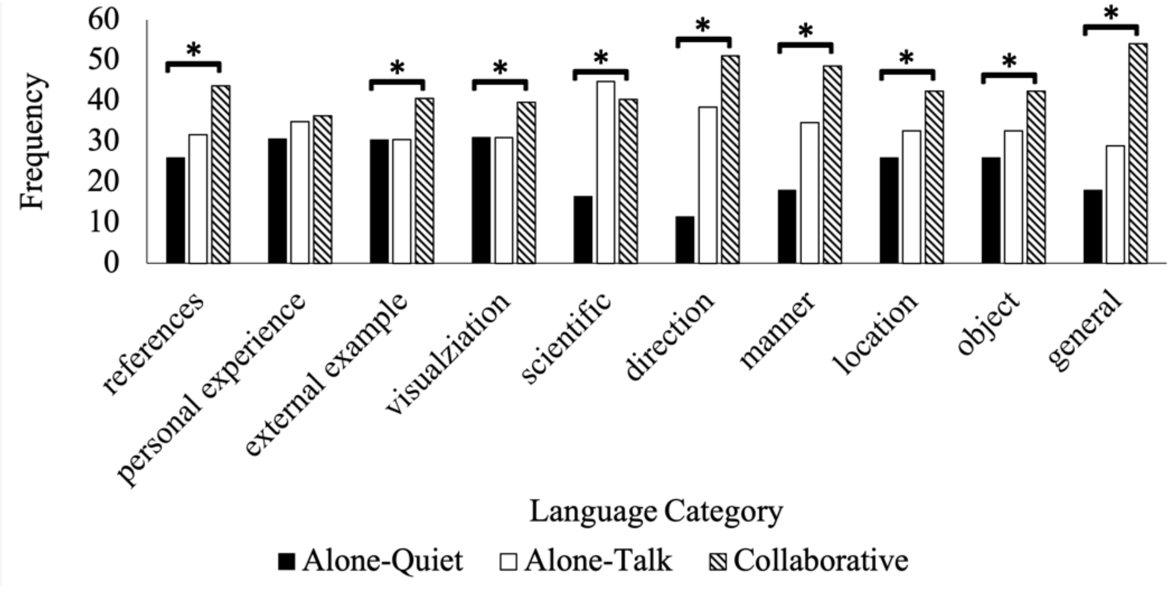

Figure 2. Frequency of utterances for each language category (excluding the "unintelligible/other" category) in each practice condition.

To further explore the content of spatial language, a MANOVA was conducted to examine condition-related differences on the frequency with which each of the ten language categories (“unintelligible/other” category was removed from analysis) were uttered (see Figure 3). The MANOVA showed that practice condition influenced the number of utterances in each category, Wilk’s Lambda = 3.01, $F(22,108) = 8.14$, $p < .01$, $\eta_p^2 = .63$. However, Levene’s test of equality of error variances was significant for all the language categories (*p*s = .00 - .02), indicating that the variance within each language category was not equal. Therefore, Kruskal-Wallis non-parametric tests were used to confirm the results of the MANOVA. As shown in Table 5, the Kruskal-Wallis H tests confirmed condition-related differences for ten out of the eleven language categories.

Finally, we examined the relationship between spatial language and accuracy on the kinematic practice problem set. Pearson’s correlations indicated no significant relationships between practice problem scores and the total number of utterances (including the unintelligible/unrelated category: $r = .15$, $p = .23$, excluding the unintelligible/unrelated category: $r = .15$, $p = .22$) or the number of utterances in any of the 11 categories (*r*s ranging from .05 to .21, *p*s ranging from .08 to .67).

Table 5

Kruskal-Wallis H Tests for Language Categories

Utterance Category	χ^2	df	Asymp. Sig.	Condition	Mean Rank
references to other problems in the problem set	17.66	2	.000*	Alone – Quiet	26.00
				Alone – Talk	31.75
				Collaborative	43.80
personal experience	2.70	2	.260	Alone – Quiet	30.64
				Alone – Talk	34.93
				Collaborative	36.33
external example	14.67	2	.001*	Alone – Quiet	30.50
				Alone – Talk	30.50
				Collaborative	40.70
visualization prompt	12.39	2	.002*	Alone – Quiet	31.00
				Alone – Talk	31.00
				Collaborative	39.74
scientific/physics related concepts	30.48	2	.000*	Alone – Quiet	16.50
				Alone – Talk	44.82
				Collaborative	40.39
direction of motion	50.41	2	.000*	Alone – Quiet	11.50
				Alone – Talk	38.55
				Collaborative	51.17
manner of motion	32.68	2	.000*	Alone – Quiet	18.00
				Alone – Talk	34.66
				Collaborative	48.67
location in space	14.76	2	.001*	Alone – Quiet	26.00
				Alone – Talk	32.64
				Collaborative	42.48
object description	14.57	2	.001*	Alone – Quiet	26.05
				Alone – Talk	32.64
				Collaborative	42.48
general movement	47.70	2	.000*	Alone – Quiet	18.00
				Alone – Talk	29.00
				Collaborative	54.09
unintelligible/other	46.25	2	.000*	Alone – Quiet	14.64
				Alone – Talk	32.45
				Collaborative	54.00

*$p < .01$.

Near and Far Transfer Assessments

Of the three near transfer problems, participants averaged 0.50 correct responses (*SD* = .62, mean proportion = .17), which is similar to findings from past studies comparing familiar and abstract problem performance (e.g., Kaiser, Jonides, & Alexander, 1986). Of the 10 paper-folding far transfer problems, participants averaged 4.5 correct responses (*SD* = 1.94, mean proportion = .45) and of the 21 cube comparison far transfer problems, participants averaged 14.90 correct responses (*SD* = 4.15, mean proportion = .71). Due to better performance on the cube comparison task, the two transfer tasks were kept separate during subsequent analyses.

Again, contrary to our hypothesis, participants who practiced with a peer did not perform better on transfer tasks compared to participants who practiced alone (see Figure 4). Considering each transfer task had a different number of problems, we standardized the accuracy score for each task by dividing the number of correct responses by the number of completed responses. Using the proportion of correct responses to completed responses as the dependent variable, we conducted a 3 X 3 ANCOVA to assess the effect of condition (Alone-Quiet, Alone-Talk, Collaborative) and transfer task (kinematic, paper-folding, cube comparison). The covariates were gender, duration of practice, physics education, spatial experience, and spatial ability. Due to a violation of assumption of sphericity, $\chi^2(2) = 10.87$, $p < .01$, a Huynh-Feldt correction was used ($\varepsilon = 0.99$). There was no main effect of practice condition, $F(2, 81) = 2.94$, $p = .06$, $\eta_p^2 = .07$, or an interaction between condition and transfer task, $F(3.97, 160.91) = 2.22$, $p = .07$, $\eta_p^2 = .05$. There was only a main effect of transfer task, $F(1.99, 160.91) = 9.12$, $p < .000$, $\eta_p^2 = .10$, due to different performances between all three tasks. The mean score was highest for the far transfer cube comparison task ($M = 0.87$, $SD = 0.14$), followed by the far transfer paper-folding task ($M = 0.69$, $SD = 0.26$), then by the near transfer kinematic task ($M = 0.17$, $SD = 0.21$). As seen in

Table 6, post hoc tests using Bonferroni correction indicated that performance on transfer tasks was significantly different between all tasks (*p*s < .01, *d* > .90).

Table 6

Bonferroni Pairwise Comparisons for Transfer Task of Accuracy

Comparisons	Mean Accuracy Difference	Std. Error	95% CI Lower Bound	95% CI Upper Bound
Kinematic Near Transfer vs. Paper-folding Far Transfer	-0.51*	0.03	-0.59	-0.43
Kinematic Near Transfer vs. Cube Comparison Far Transfer	-0.71*	0.03	-0.77	-0.65
Paper-folding Far Transfer vs. Cube Comparison Far Transfer	-0.20*	0.03	-0.27	-0.13

$^{*}p < 0.01$

Lastly, although gender ($p = .24$, $\eta_p^2 = .02$), duration of practice ($p = .57$, $\eta_p^2 = .00$) physics education ($p = .25$, $\eta_p^2 = .02$) and spatial experience ($p = .93$, $\eta_p^2 = .00$) were not significant predictors, individual spatial ability was a significant predictor of accuracy on transfer tasks ($p = .00$, $\eta_p^2 = .13$). The relationship between individual spatial ability and transfer task accuracy was assessed using a Pearson's correlation (see Figure 5 and 6). Higher individual spatial ability was positively correlated with accuracy on both the far transfer paper-folding task ($r = .43$, $p < .01$) and the far transfer cube comparison task ($r = .28$, $p < .01$).

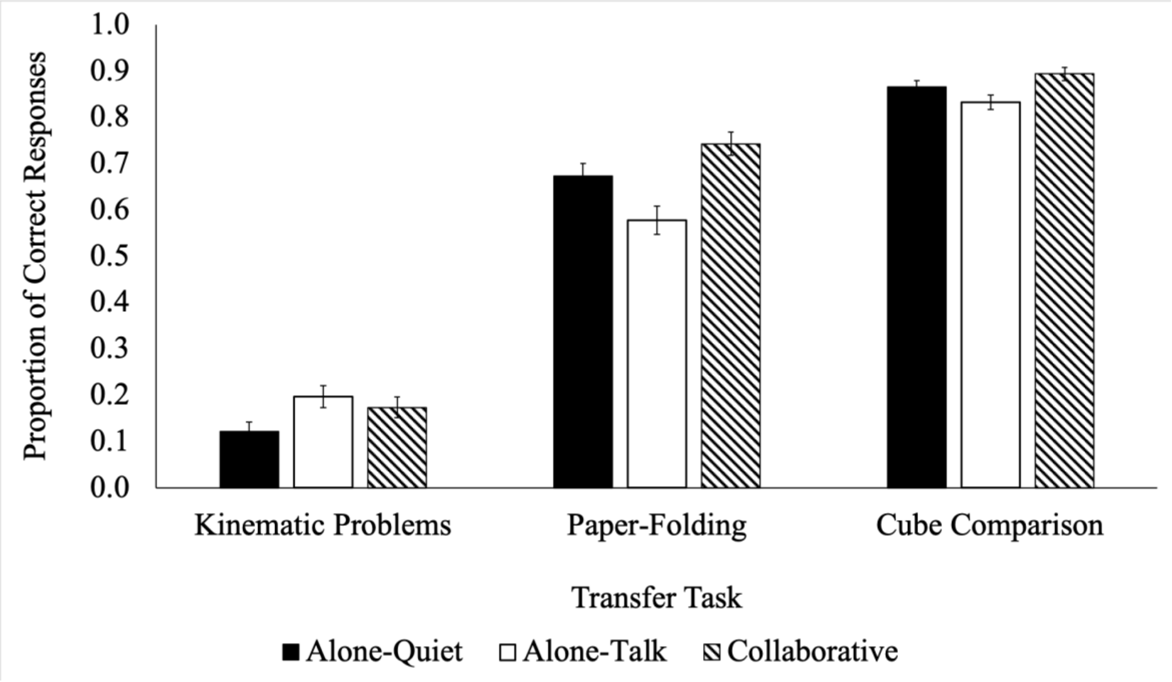

Figure 4. Accuracy on transfer tasks for each practice condition. Error bars indicate mean standard error.

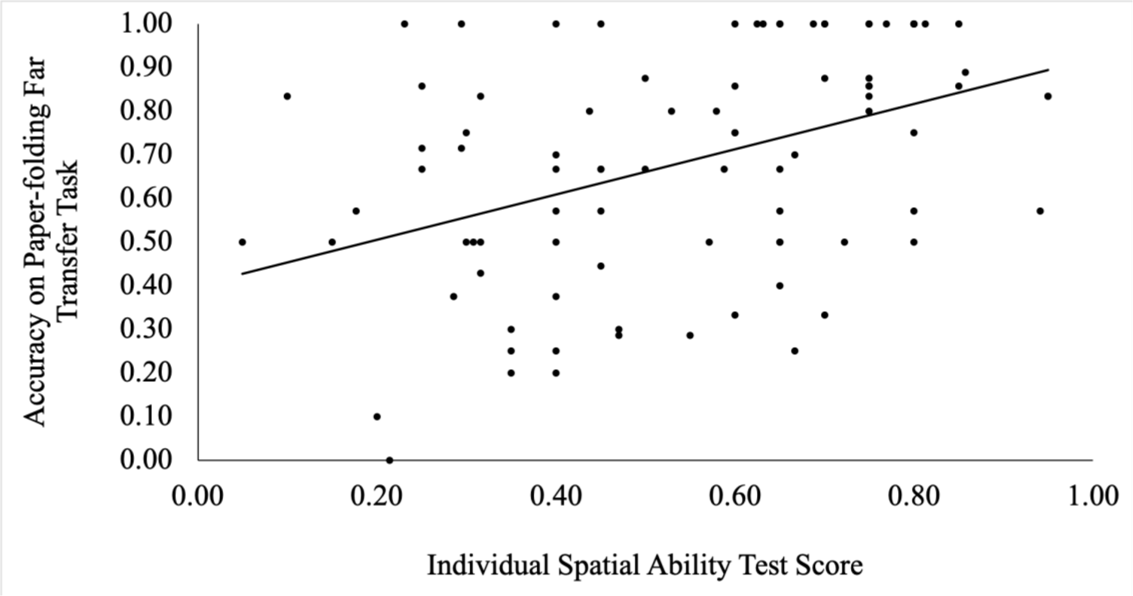

Figure 5. A Pearson's correlation between individual spatial ability test scores and accuracy on the far transfer paper-folding task. Individual spatial ability test scores are the proportion of correct responses to completed problems.

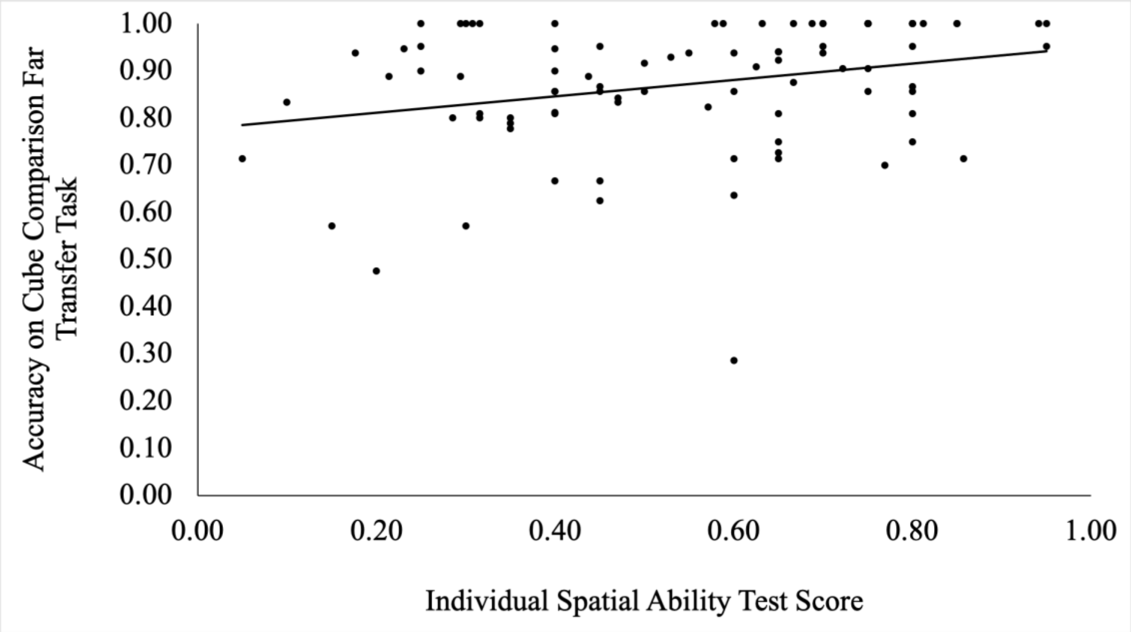

Figure 6. A Pearson's correlation between individual spatial ability test scores and accuracy on the far transfer paper-folding task. Individual spatial ability test scores are the proportion of correct responses to completed problems.

We wondered whether participants in the Collaborative condition would perform better as a result of more task-related discussion, which may facilitate greater spatial reasoning. A Pearson's correlation revealed a positive relationship between the total number of utterances by an individual and the proportion of correct responses on the near transfer kinematic assessment ($r = .23$, $p = .03$), but no significant relationship between the utterances and accuracy on the paper-folding task ($r = .15$, $p = .15$) or cube comparison task ($r = -.03$, $p = .76$). When we removed unintelligible/unrelated language, the correlation between the total amount of utterances remained significant for the near transfer kinematic problems ($r = .23$, $p = .03$) and remained not significant for the paper-folding task ($r = .09$, $p = .40$) and cube comparison task ($r = -.05$, $p = .63$). More specifically, the use of visualization prompts ($r = .21$, $p < .05$), location words/phrases ($r = .24$, $p = .02$), and object descriptions ($r = .21$, $p < .05$) were related to higher scores on the near transfer kinematic assessment.

Overall, peer collaboration did not affect accuracy on related kinematic and spatial tasks. However, results indicate that specific types of language were positively correlated with success on kinematic tasks, and individual spatial ability was positively correlated with spatial tasks.

Discussion

The goal of this study was to examine whether practicing kinematic motion problems with a peer could improve kinematic problem solving and facilitate greater spatial reasoning. We reasoned that working with a partner would lessen the cognitive load related to mentally envisioning an object motion event while simultaneously enabling an individual to receive feedback about his or her ideas. To our surprise, we found that participants in the Collaborative condition did not perform any better on the practice problems, near transfer problems, or far transfer problems compared to participants in the Alone conditions. Additionally, we did not find duration of practice, physics education, or spatial experience to be significant predictors of accuracy on the problem sets, even though previous studies have found these variables to positively influence kinematic problem solving. One explanation for why the covariates were not significant is the low variability within the variables. However, we did expect to find a positive relationship between spatial ability and performance on far transfer tasks, considering the paper-folding and cube comparison problems required a large amount of mental transformation. But, the lack of relationship between spatial ability and near transfer kinematic problems was unexpected, suggesting that even though kinematic problems require a spatial ability, spatial ability alone may not be enough to facilitate kinematic problem solving.

A possible explanation for this (null) finding is that the practice session only occurred once and for a few minutes. In previous studies, successful findings were related to practice sessions taking place over the course of a semester or academic year (Azmitia, 1988; Dimant & Bearison, 1991; Fawcett & Garton, 2005; Kozhevnikov & Thornton, 2006; Phelps & Damon, 1989). Practice sessions that occurred only once tended to have lower rates of success (Golbeck & Sinagra, 2000). Increasing the number of practice sessions may allow individuals more time to contemplate the materials and think of new ideas to share at a later session as well as enable individuals to build a relationship with their partner in order to feel comfortable sharing their ideas and offering feedback.

By extension, increasing the number of practice sessions may serve to improve the quality of discussion between partners. We found duration of practice to be a significant predictor for the amonut of language and amount of spatial language used during the practice portion, suggesting that if participant had more time they would generate more task-related and spatial language. Although, it is important to recognize that even though participants in the Collaborative condition spoke more frequently and used more spatial and task-related language on average (e.g., direction of motion, manner of motion, references to other problems in the problem set) than participants in the Alone conditions, collaborators were often only parroting what their partner said or mimiking the language in the directions. Indeed, 59% of all utterances in the collaborative condition were codeed as “unintellible/other.” Additionally, participants sometimes appear to lack deep conceptual understanding and did not follow their ideas up with explanations. For example, participants would read the directions, offer a prediction (e.g., “[the ball] will go straight down”), and then support their prediction by using a word or idea from the directions (e.g., “because there is no air resistance”) without explaining what that means. Similarly, participants often used a science-related word (e.g., gravity, velicity) to support their prediction but did not offer further explanation about how gravity or velicity worked to influence the object’s motion. Therefore, it would be beneficial to divide our language categories further, parsing apart mentions of concepts and explanations using conceptual words, and to systematically vary the duration of time participants spend working on the practice problems.

Another explanation for the nonsignificant results is that paper-and-pencil tasks are not conducive for learning kinematic concepts and generating conversation between partners. When a task relies on an internal representation, it is difficult to present that representation to another person and identify its strenghts and weaknesses. For example, the Piagian Water Level task relied on individuals' kinesthetic sense of horizontality and made it difficult for collaborators to discuss the task. Therefore, partners resorted to confirming their partner's ideas rather than critiquing them (Golbeck & Sinagra, 2000). Perhaps using 3-D models or having collaborators act out object motion events would facilitate better discussion. In studies with younger children, spatial tasks often involve physical toys (e.g., Legos), making it easier to demonstrate what they think will happen or see how one object influences another (Azmitia, 1988; Fawcett & Garton, 2005). Having objects to manipulate and an additional person to observe the object motion event might be the key to dispelling misinformation gathered from sensory input, as suggested by the actions-on-objects hypothesis, or from perceptual illusions, as suggested by the seeing-is-believing hypothesis, once partners compare perspectives of how the object moved.

The use of 3-D models or toys may also enable participates to generate more spatial language on their own. Using objects when asking participants to predict where and how an object will move would eliminate the need for detailed directions with descriptive language regarding the object, how it is positioned, and how it is initally moving because the participants would be able to see and feel the object, and witness how the object is positioned and moving by either directly manipulating the object or watching an experimenter manipulate the object. Also, we would expect that using manipulables would encourage participants to interact with the object as opposed to strictly relying on the description in the directions. Having an object to directly interact with might lessen the cognitive load asociated with holding a mental representation of the object's movement, and allow participants to consider *why* the object is moving in a specific way and offer more substantial explanations.

We did find a positive correlation between the number of utterances by an individual and proportion of correct responses on the near transfer kinematic problem set. More specifically, the use of visualization prompts, location words/phrases, and object descriptions were directly correlated with higher scores on the near transfer kinematic assessment. This could suggest that using words or phrases related to these categories is promoting greater spatial reasoning because the participant is considering where the object is relative to other objects (i.e., establishing a frame of reference with which to compare the object's movement, as suggested by the see-is-believing hypothesis), how the object feels and looks (i.e., thinking about the allowances of its shape and size, as suggested by the actions-on-object hypothesis), and either prompting their partner or themselves to imagine some aspect of the problem (i.e., discussing a mental representation of the event). Although, it is also possible that participants who are using these types of language are better spatial reasoners and, therefore, have an easier time generating spatially rich language. However, we have to be cautious when interpreting these findings because the interrater reliability rates for these categories are either less than .50 or were unable to be calculated because there were too few cases.

Previous studies have shown that the use of spatial language is important to promote spatial reasoning (e.g., Pruden, Levine, & Huttenlocher, 2011). We developed a new coding scheme based on previous coding manuals for spatial tasks (Cannon, Levine & Huttenlocher, 2007; Teasley, 1995; Winsler, Fernyhough, McClaren & Way, 2005). We focused on categories that related to specific words or phrases that conveyed a spatial concept or an idea related to the assessments. However, measuring the use of spatial language might not be the most effective way to measure the quality of discussion between partners. When individuals are collaborating, there is more to just the content of what they are communicating. The language they use has a function, such as asking a question for clarification or making a prediction. The coding scheme we used for this study did not indicate the function of the language. Therefore, a question using the word "straight-down" was coded the same way as a prediction using the word "straight-down". This coding scheme failed to capture the complete picture behind the content being used. Therefore, a coding scheme that focuses on content and function (i.e., the purpose of an utterance) might be a more accurate way to assess the quality of discussion. Additionally, the coding scheme did not capture the nuances of the discussion, such as how often one partner was confirming the ideas of the other partner versus critiquing their ideas. Golbeck & Sinagra (2000) noted that women often used more confirmatory language when working with a peer compared to men, and women used more confirmatory language when working with a male partner as opposed to a female partner. This type of dyadic interaction may limit the effectiveness of working with a partner if one of the partners does not contribute equally. Therefore, in addition to an improved coding scheme, another method to assess quality of discussion could be adminstering follow-up questionnaires to ask participants whether they think the discussion with their partner was constructive and if they agreed with all the given answers. Receiving feedback from the participants could help uncover which aspects of talking with a peer faciliate problem solving. Overall, our coding scheme did not account for all the possible ways to assess quality of discussion. Therefore, our results should be interpretted with caution because they focus on a specific element of the dyadic interactions.

Additionally, we utilized a novel coding software that allowed us to develop our own coding procedure based on the purpose of the study. Since we were interested in the frequency and content of the utterance, but not the duration, we did not account for temporal specificity when marking the occurrence of language, meaning coders were asked to indicate when an utterance occurred but not whether to mark it at the beginning, middle, or end of the utterance. This discrepancy between coders could have contributed to the low inter-rater reliability rates reported for some of the language categories. Additionally, we coded the video recording in real-time and did not transcribe the audio prior to coding for language. Therefore, coders may have misheard or did not hear portions of the video, contributing to the low inter-rater reliability rates.

Overall, the findings of this study are important to consider and have educational implications. We found that practicing with a partner was not an effective way to improve kinematic problem solving or related spatial abilities. However, we should consider the limitations of the task modality, the duration of practice, and the quality of discussion between partners when evaluating the effectiveness of peer collaboration. Future studies should consider incorporating: interactive models (to provide important visual and haptic feedback as well as enable peers to easily discuss specific task components and solutions), recurring practice sessions (to enable peers multiple opportunities to work together and build a rapport), and multidimensional language categories (to capture the content, context, and purpose of the utterances). Lastly, future studies should aim to evaluate the efficacy of peer collaboration and its effect on kinematic problem solving.

References

Azmitia, M. (1988). Peer interaction and problem solving: When are two heads better than one? *Child Development*, *59*(1), 87–96. https://doi.org/10.1111/1467-8624.ep10514042

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*(5–6), 327–344. https://doi.org/10.1007/BF0028772

Cannon, J., Levine, S., & Huttenlocher, J. (2007). *A system for analyzing children and caregivers' language about space in structured and unstructured contexts*. Spatial Intelligence and Learning Center (SILC) technical report. Retrieved from *http://moustachio.cs.northwestern.edu/media/silc_pdfs/resources/ testsandinstruments/Spatial%20Language%20Coding%20Manual%201-10-12.pdf*.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition*, *9*(2), 117–123. https://doi.org/10.1016/0010-0277(81)90007-X

Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics*, *48*(12), 1074–1079. https://doi.org/10.1119/1.12290

Chatterjee, A. (2008). The neural organization of spatial thought and language (Vol. 29, pp. 226–238). Presented at the Seminars in Speech and Language, Thieme Medical Publishers. https://doi.org/10.1055/s-0028-1082886

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*(1), 66–71. https://doi.org/10.1119/1.12989

Cook, N. J., & Breedin, S. D. (1994). Constructing naive theories of motion on the fly. *Memory & Cognition*, *22*(4), 474–493. https://doi.org/10.3758/BF03200871

Dancu, T., Gutwill, J., & Sindorf, L. (2009). Geometry Playground Pathways Study. *San Francisco: Exploratorium Science Museum*. *Hosted by the Exploratorium Science Museum, http://exploratorium.edu/vre/pdf/Geometry_Playground_Pathways.pdf.*

Dancu, T., Gutwill, J., & Sindorf, L. (2015). Geometry Playground Spatial Language Coding Manual. Retrieved from *http://spatiallearning.org/media/silc_pdfs/resources/ testsandinstruments/GP_Spat_Reas_scheme_forSILC_11. pdf*.

Datavyu Team. (2014). Datavyu: A Video Coding Tool (Version v1.4.1). New York University: Datavyu Project. Retrieved from http://datavyu.org.

Dimant, R. J., & Bearison, D. J. (1991). Development of formal reasoning during successive peer interactions. *Developmental Psychology*, *27*(2), 277–284. https://doi.org/10.1037/0012-1649.27.2.277

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests* (Vol. 102). Educational Testing Service Princeton, NJ.

Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem solving ability. *British Journal of Educational Psychology*, *75*(2), 157–169. https://doi.org/10.1348/000709904X23411

Golbeck, S. L., & Sinagra, K. (2000). Effects of gender and collaboration on college students' performance on a Piagetian spatial task. *Journal of Experimental Education*, *69*(1), 22–35. https://doi.org/10.1080/00220970009600647

Hogan, D. M., & Tudge, J. R. H. (1999). Implications of Vygotsky's theory for peer learning. In A. M. O'Donnell, A. King, A. M. O'Donnell (Ed), & A. King (Ed) (Eds.), *Cognitive perspectives on peer learning.* (pp. 39–65). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=psyh&AN=1999-02359-002&site=ehost-live&authtype=sso&custid=s8475574

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. *The Analogical Mind: Perspectives from Cognitive Science*, 1–19. Retrieved from http://reasoninglab.psych.ucla.edu/wp-content/uploads/2010/09/The-Place-of-Analogy-in-Cognition.pdf

Isaak, M. I., & Just, M. A. (1995). Constraints on the processing of rolling motion: The curtate cycloid illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1391. http://dx.doi.org/10.1037/0096-1523.21.6.1391

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137. http://dx.doi.org.proxy1.library.jhu.edu/10.1037/0033-295X.92.2.137

Kaiser, M., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, *14*(4), 308–312. https://doi.org/10.3758/BF03202508

Kozhevnikov, M., & Garcia, A. (2011). Visual-spatial learning and training in collaborative design in virtual environments. In *Collaborative design in virtual environments* (pp. 17–26). Springer. https://doi.org/10.1007/978-94-007-0605-7_2

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*(3), 439–453. https://doi.org/10.3758/BF03196179

Kozhevnikov, M., Hegarty, M., & Mayer, R. (2002). Spatial abilities in problem solving in kinematics. In *Diagrammatic representation and reasoning* (pp. 155–171). Springer. https://doi.org/10.1007/978-1-4471-0109-3_9

Kozhevnikov, M., Hegarty, M., & Mayer, R. (1999). Students' use of imagery in solving qualitative problems in kinematics. Marion, IN: Indiana Wesleyan Center for Educational Excellence. Retrieved from ERIC database. (ED433239)

Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education and Technology*, *15*(1), 111. https://doi.org/10.1007/s10956-006-0361-0

Kozhevnikov, M., Motes, M., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*(4), 549–579. https://doi.org/10.1080/15326900701399897

Landau, B., & Jackendoff, R. (1993). Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences*, *16*(2), 255–265. https://doi.org/10.1017/S0140525X00029927

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental psychology*, *48*(2), 530. http://dx.doi.org.proxy1.library.jhu.edu/10.1037/a0025913

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479–1498. https://doi.org/10.2307/1130467

McCloskey, M. (1983a). Intuitive physics. *Scientific American*, *248*(4), 122–130. Retrieved from https://s3.amazonaws.com/academia.edu.documents/48512572/scientificamerican0483-12220160902-7181-1vws9g6.pdf?response-contentdisposition=inline%3B%20filename%3DIntuitive_Physics.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190731%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190731T155020Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=060de3cf3ba14b79debeef398eb4d484e045b46dc649ed1e0931336223238d68

McCloskey, M. (1983b). Naive theories of motion. In D. Genter & A. L. Stevens (Eds.), *Mental Models* (pp. 299–324). Hillsdale, NJ: Lawrence Erlbaum Associates.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science*, *210*(4474), 1139–1141. https://doi.org/10.1126/science.210.4474.1139

McCloskey, M., & Kaiser, M. (1984). The impetus impulse. *The Sciences*, *24*(6), 40–46. https://doi.org/10.1002/j.2326-1951.1984.tb02754.x

McCloskey, M., & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 146–156. https://doi.org/10.1037/0278-7393.9.1.146

McCloskey, M., Valtonen, J., & Cohen Sherman, J. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680–713. https://doi.org/10.1080/02643290500538356

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 636–649. https://doi.org/10.1037/0278-7393.9.4.636

Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology*, *81*(4), 639–646. https://doi.org/10.1037/0022-0663.81.4.639

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter?. *Developmental science*, *14*(6), 1417-1430. https://doi.org/10.1111/j.1467-7687.2011.01088.x

Salthouse, T. A., Babcock, R. L., Mitchell, D. R., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence*, *14*(2), 187–230. https://doi.org/10.1016/0160-2896(90)90004-D

Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R. D., & Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, *26*(1), 128–136. https://doi.org/10.1037/0012-1649.26.1.128

Salthouse, T. A., Mitchell, D. R., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 507–516. https://doi.org/10.1037/0278-7393.15.3.507

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, *31*(2), 207. http://dx.doi.org/10.1037/0012-1649.31.2.207

Tudge, J. (1992). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 155-172). New York, NY, US: Cambridge University Press.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Vygotsky, L. (1978). Interaction between learning and development. In Gauvian & Cole (Eds.), *Readings on the Development of Children* (pp. 34-40). New York: Scientific American Books.

Vygotsky, L. S., & Cole, M. (1978). *Mind in Society : Development of Higher Psychological Processes*. Cambridge: Harvard University Press. Retrieved from http://search.ebscohost.com.proxy1.library.jhu.edu/login.aspx?direct=true&db=nlebk&AN=575543&site=ehost-live&scope=site

White, P. A. (2009). Property transmission: An explanatory account of the role of similarity information in causal inference. *Psychological Bulletin*, *135*(5), 774. http://dx.doi.org/10.1037/a0016970

White, P. A. (2012). The impetus theory in judgments about object motion: A new perspective. *Psychonomic Bulletin & Review*, *19*(6), 1007–1028. https://doi.org/10.3758/s13423-012-0302-2

White, P. A. (2013). Judgements about the relation between force and trajectory variables in verbally described ballistic projectile motion. *Quarterly Journal of Experimental Psychology*, *66*(5), 876–894. https://doi.org/10.1080/17470218.2012.721790

Winsler, A., Fernyhough, C., McClaren, E.M., & Way, E. (2005). Private speech coding manual. Unpublished manuscript. George Mason University, Fairfax, VA, USA. Retrieved from http://classweb.gmu.edu/awinsler/Resources/PsCodingManual.pdf